Supercell-size convergence of formation energies and gap levels of vacancy complexes in crystalline silicon in density functional theory calculations

J. Dabrowski* and G. Kissinger
IHP, Im Technologiepark 25, 15236 Frankfurt (Oder), Germany
(Received 30 October 2014; revised manuscript received 20 April 2015; published 12 October 2015)

Results for supercell-size convergence of formation energies and charge transition levels of vacancy complexes V_n ($1 \leqslant n \leqslant 11$) in crystalline Si are reported for the *ab initio* density functional theory (DFT) with generalized gradient approximation (GGA) pseudopotentials. When extrapolated to the dilute limit, the formation energy of an uncharged vacancy becomes 3.74 ± 0.1 eV, and the binding energy of an uncharged divacancy becomes 1.9 ± 0.2 eV. Stable V_n clusters are built on the basis of sixfold rings ($n \leqslant 6$) and of octahedral voids ($n \geqslant 7$). If the well-known underestimate of the band gap by the DFT and the accuracy of extrapolations are taken into account, the extrapolated levels are in good agreement with experiment. We discuss the implications for simulation of vacancy clustering during thermal quenching, for interpretation of deep-level spectroscopy and electron paramagnetic resonance in irradiated Si, and for cell and Brillouin-zone sampling choice when DFT-related methods beyond local-density approximation or GGA are used.

DOI: 10.1103/PhysRevB.92.144104 PACS number(s): 61.72.Bb, 61.72.jd, 61.72.uf, 63.20.dk

I. INTRODUCTION

A single silicon vacancy V_1 is probably the most studied and best understood defect in crystalline Si. It has been thoroughly characterized experimentally [1–8] and theoretically [9–30]. This simple intrinsic point defect exists in low concentrations (below 10^{14} cm^{-3} even at 800 °C) at thermal equilibrium with the perfect crystal at usual conditions [7] but is formed in significant amounts by thermal quenching from well above 1000 °C, by particle irradiation, during some reactions of the crystal with the environment (e.g., by thermal nitridation) [31], or during some defect-defect reactions. It is electrically active, diffuses fast [8,16], and mediates diffusion of some dopants [32–34]. It can be detected and identified [35] by popular methods, like deep-level spectroscopy (DLTS), electron paramagnetic resonance (EPR), infrared spectroscopy (IR), or positron annihilation spectroscopy (PAS). Its influence may also be deduced from electrical data, e.g., from carrier lifetime measurements or spreading resistance profiling (SRP). And it makes an illustrative case of Jahn-Teller distortions [4,15,20,36], being the first proven example of negative Hubbard U [4,15,18,30,37]. It combines with impurities [29], forming complexes like the A center [38] (with O) or the E center [39] (with P), and also with other defects, including other vacancies. Si vacancies and their agglomerates must thus be dealt with in the simulation of various defect-related processes, such as dopant implantation, rapid thermal annealing (RTA), and SiO_2 precipitation. Hence, multivacancies have also been widely studied [1–3,40–68] by experiment and by theory. Theoretical analysis is helpful for interpretation of DLTS [56] and EPR [20] spectra or for simulation of impurity nucleation and precipitation [67], to name just a few examples.

Since the number of atoms that can be treated by a first-principles density functional theory (DFT) calculation is quite limited, the presence of the "infinite" crystal must be accounted for in an approximate way. In semiconductor physics, the most popular solution is to impose periodic boundary conditions (PBC) and use the plane-wave formalism in the resulting periodic repetition of the atomic cluster ("supercell"). This is easy to implement and employs mature numerical procedures, and its accuracy due to the finite basis set can be tested systematically. This method has been widely used also for V_1 and its clusters V_n [20–30,49–67]. Yet it suffers from spurious effects caused by interaction between defects in the superlattice produced by the PBC [69]. The raw computed total energy must be corrected for this interaction [70–81].

The simplest alternative is to saturate the surface of the cluster with hydrogen atoms [19,54,65,82]. The price is that other finite-size effects appear: crystal bands change into discrete states, quantum localization steps in, and the relaxation field must be truncated at the cluster boundary. In addition, the influence of the H termination on the charge distribution in the cluster is hard to estimate.

One may also embed the cluster in the perfect crystal by employing Green's functions, so that the defect wave function may extend from the perturbed region into the perfect region [9–13]. This is powerful and was successfully applied to V_1 [9–18] yet is not free from drawbacks. The calculations converge slowly in energy: Green's function at energy E involves a summation of $1/[E - E_n(k)]$ over the host crystal bands $E_n(k)$. This imposes additional practical limits on the cluster size. Furthermore, truncation of all interactions at the cluster boundary makes the calculated total energy depend on the cluster size [18].

The computed V_n formation energies reported in the literature are thus to a certain extent ambiguous, and the available data are only approximately compatible with experimental observations. In particular, there are only a few reports on the *ab initio* energies of the complete series of vacancy clusters in the n range that is of interest for the simulation of such processes as SiO_2 precipitation [64–67,83].

We have thus reconsidered the problem by performing standard pseudopotential DFT calculations within the generalized gradient approximation (GGA), using periodic supercells with up to 1458 atoms. We focus on analysis of convergence with respect to the supercell size. The V_1, V_2, and V_3 defects,

*Corresponding author: Dabrowski@ihp-microelectronics.com

which have already been clearly identified by EPR and DLTS, are given the full treatment, and the results are used to support accuracy assessments. Big V_n clusters ($7 \leqslant n \leqslant 11$) are discussed mostly from the perspective of the prediction of the stable charge state for midgap Fermi levels, as the formation energies of defects in such states have implications for clustering of vacancies obtained by thermal quenching [67]. This happens at elevated temperatures, when the Fermi energy for usual doping levels approaches the midgap value [84]. The formation energies and gap states computed by us for the stable and (selected) metastable representatives of each size in the V_n series may be helpful in the interpretation of DLTS and EPR spectra in irradiated silicon.

Section II describes the approach. Section III focuses on finite cells and on benchmarking. Section IV is about the dilute limit, and Sec. V is on the implications for simulation of vacancy clustering after thermal quenching, for the interpretation of DLTS and EPR spectra in irradiated Si, and for the choice of Brillouin-zone sampling and superlattice in computationally demanding calculations.

II. APPROACH

Ab initio DFT calculations were carried out with the pseudopotential plane-wave code QUANTUM ESPRESSO (QE) [85]. Supercells containing up to 1458 Si sites were used. For exchange and correlation energy we employed the GGA in Perdew-Burke-Ernzerhof (PBE) form [86], and we allowed for spin polarization of electrons. We used the ultrasoft PBE pseudopotentials provided with QE (SI.PBE-N-VAN) and two custom PBE pseudopotentials (SI.PBE-5170 and SI.PBE-5133) [87]. The energy cutoff $E_{\rm cut}^{\rm wfc}$ was 20 Ry for the standard and 10 Ry for the custom pseudopotentials. The bulk properties of Si (lattice constant and bulk modulus) are described with typical DFT accuracy [88,89]. SI.PBE-5170 was constructed to reproduce at lower costs ($E_{\rm cut}^{\rm wfc} = 10\,{\rm Ry}$) the bulk properties of Si obtained [67] by SI.PBE-N-VAN ($E_{\rm cut}^{\rm wfc} = 20\,{\rm Ry}$), while SI.PBE-5133 was constructed to reproduce the experimental bulk properties of Si as closely as possible at low costs ($E_{\rm cut}^{\rm wfc} = 10\,{\rm Ry}$); see Table I. No significant effect of the pseudopotential choice on V_n^q formation energies was observed.

Formation energies are defined with respect to the Si bulk: the removed atoms are transferred to a reservoir in which they have the same energy as in the bulk. If the defect is charged, the charge comes from/to the Fermi level $E_{\rm F}$. The formation energy $E_{\rm form}(n,q,E_{\rm F})$ of a V_n^q cluster computed in an m-atomic supercell is therefore

$$E_{\rm form}(n,q,E_{\rm F}) = E_{\rm tot}^{(m)}(n) - \frac{m-n}{m}E_{\rm tot}^{(m)} + qE_{\rm F}, \qquad (1)$$

TABLE I. lattice constant (Å), bulk modulus (GPa), and band gap of Si (eV). Experimental data [84,90,91] compared to values obtained by GGA DFT pseudopotentials used in this work [87].

	Lattice constant	Bulk modulus	Gap
Experiment	5.430 [90]	97.6 [91]	1.12 [84]
SI.PBE-5133	5.433 (0.05%)	94 (−3.7%)	0.76
SI.PBE-5170	5.472 (0.77%)	90 (−7.8%)	0.76
SI.PBE-N-VAN	5.469 (0.71%)	88 (−9.8%)	0.76

where $E_{\rm tot}^{(m)}(n)$ and $E_{\rm tot}^{(m)}$ are the computed total energies of V_n and of perfect Si. The dissociation energy $D(n,q_n,E_{\rm F})$ of vacancy from the cluster $V_n^{q_n}$ is the energy lost in the reaction $V_n^{q_n} \rightarrow V_1^{q_1} + V_{n-1}^{q_{n-1}}$; hence,

$$\begin{aligned} D(n,q,E_{\rm F}) = {} & E_{\rm form}(n-1,q_{n-1},0) \\ & + E_{\rm form}(1,q_1,0) - E_{\rm form}(n,q,0) \\ & + (q_{n-1}+q_1-q_n)E_{\rm F}. \end{aligned} \qquad (2)$$

Gap levels (q_1/q_2) of a defect correspond to Fermi energies at which formation energies of charge states q_1 and q_2 of the defect are the same. Their location in the gap is therefore affected not only by the energies of the defect orbitals but also by relaxation of atoms in the different charge states. These charge transition levels can be found by plotting the formation energies given by Eq. (1) against the Fermi energy $E_{\rm F}$ or directly from the formula

$$(q_1/q_2) = \frac{E_{\rm form}(q_2,E_{\rm VB}) - E_{\rm form}(q_1,E_{\rm VB})}{q_1-q_2} - E_{\rm VB}, \qquad (3)$$

where $E_{\rm VB}$ is the energy of the DFT valence-band maximum (VBM). The latter can be obtained by calculating the total energy of the perfect crystal in the neutral charge state and with one hole in the valence band,

$$E_{\rm VB} = E_{\rm tot}^{({\rm perfect})(0)} - E_{\rm tot}^{({\rm perfect})(+)}. \qquad (4)$$

$E_{\rm VB}$ approximately equals the DFT VBM eigenvalue [92].

GGA and local-density approximation (LDA) DFT band edges do not coincide with the true band edges. This is caused by artificial self-interaction of electrons and by the lack of discontinuity in the exchange-correlation potential derivative with respect to the occupation number [93]. A correction is needed when DFT levels are compared to experiment [77,78,94]: the former should be shifted by $dE_{\rm F}$. One may argue that $dE_{\rm F}$ should be the same for all defects [26]. Some authors use here the "marker method": $dE_{\rm F}$ is chosen so that when corrected, a selected level coincides with its established experimental position [28,53,94–96]. We consider the levels of V_1, V_2, and V_3 as marker candidates for V_n and eventually choose the V_1 and V_3 levels. The best choice is not fully defect and state independent, as it relates to how the host bands admix to the defect state.

We now turn to the choice of Bravis lattice. The practically useful range of cell sizes is between about 60 to less than 2000 atoms. Figure 1 shows the available number of atoms for cells having the full symmetry of the bulk (fcc, sc, and bcc). More flexibility is obtained when cells of lower symmetry (see Appendix A) are added to the set [79]. Using cells of various types results, however, in nonmonotonic dependence of the total energy on the number of atoms. Indeed, the quality of the Brillouin-zone (BZ) sampling varies smoothly with the number of atoms only for cells of the same type. The same is true for the interactions in the superlattice of defects. We have thus restricted the choice to a single Bravis lattice.

The size convergence is the fastest for bcc because the overlap of wave functions is the smallest [24,30] and because the alignment between the perfect bonds and the spurious charge induced by finite size is most favorable in this case [24]. As far as these aspects are considered, there is no fundamental difference between sc and fcc [24]. For a single vacancy, the sc

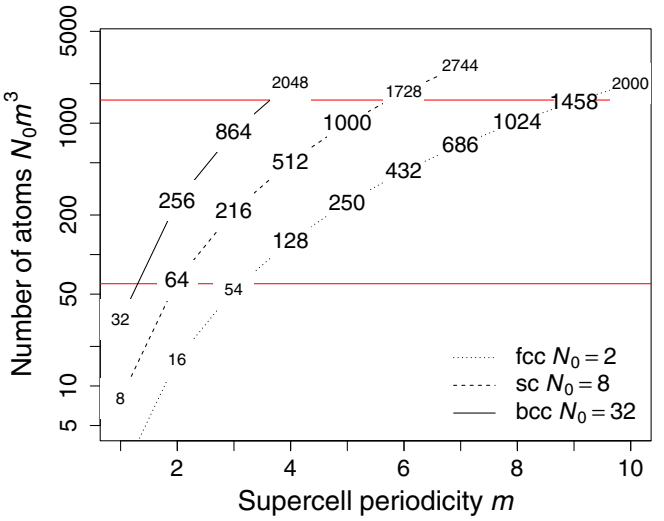

FIG. 1. (Color online) Number of atoms in fcc, sc, and bcc cells. The range useful for a convergence study is delimited by horizontal lines.

cells perform better than fcc [24,30], but this is not necessarily a general rule. Since there are more fcc than sc cells available (Fig. 1), we chose fcc cells.

The cell-size (r) variation of $E_{\rm tot}$ is smaller when computed with special k points in the BZ than when computed with only the BZ center (Γ) [21,24,30,78]. Here, the Γ scheme was chosen for three main reasons. First, we study numerous charge states of numerous defects and defect structures. In most cases we thus encounter partially filled Kohn-Sham states in the gap that are still dispersive even in large cells. If multiple k points are used, unphysical and uncontrollable effects may thus occur [30,97]. Furthermore, the energy of a defect level at Γ converges with the cell size faster than the center of weight of the defect band does [30]. Last but not least, the computing time per iteration done at Γ is about two time shorter than needed to perform the same calculation for any other k point in the BZ interior [98]. However, although for cells used here the computed lattice constant and bulk modulus are converged or nearly converged at Γ [99], small size-dependent inaccuracies in the description of the host bonds remain [99] and, when summed over the supercell, contribute to the size dependence of $E_{\rm tot}$.

The unwanted side effect of periodic boundary conditions and finite supercell size is that each defect sees a periodic superlattice of defects with which it interacts: quantum chemically (due to overlap of defect wave functions) [30,73,77,78], elastically (due to overlap of defect strain fields) [28,77,78], and electrostatically [69–78,80,81]. This spurious interaction vanishes at infinite separation between defects, but at the separation used in manageable DFT calculations it should be corrected for or at least estimated to assess the reliability of the results.

The finite-size correction has been formally analyzed in the seminal paper by Makov and Payne (MP) for the case of electrostatic interactions between superlattice of charged objects in the presence of uniform neutralizing background [71]. MP found that the dependence of total energy $E_{\rm tot}$ on the supercell dimension r is in this case

$$E_{\rm tot} = E_\infty - \frac{q^2\alpha}{2\epsilon r} - \frac{2\pi q Q}{3\epsilon r^3} + O(r^{-5}), \qquad (5)$$

where q is the charge, Q is the quadrupole moment, ϵ is the static dielectric constant, and α is the Madelung constant. If r is taken as a cube root of the supercell volume, then $\alpha_{\rm fcc} = 2.885$, $\alpha_{\rm sc} = 2.373$, and $\alpha_{\rm bcc} = 2.883$ [72].

Since the strain field $u(r)$ caused by a spherical inclusion in an isotropic medium decays as $1/r$ and the elastic energy density is proportional to u^2, the elastic and electrostatic contributions of finite cell size to $E_{\rm tot}$ have the same functional form that may be approximated by retaining the first two terms in the multipole expansion:

$$\Delta E = -|A|r^{-1} + Br^{-3}. \qquad (6)$$

The absolute value ensures that the monopole term, $-|A|r^{-1}$, is attractive, i.e., that the monopole contribution to the formation energy contributes to the increase of the total energy with increasing defect-defect separation. This requirement may be somewhat counterintuitive, given that charges of the same sign repel one another. Yet for electrostatic interactions this effective attraction is the consequence of the neutralizing background that is added in the supercell calculation to each charged cell in order to keep the cells electrically neutral [71]. This behavior is illustrated in Fig. 2 for the simple case of a negatively charged oxygen atom placed in vacuum. The sign of the second term, Br^{-3}, is equal to the sign of the product qQ, where q is the charge and Q is the quadruple moment [71].

Elastic interactions in the isotropic medium are similar in form to electrostatic interactions. In particular, the strain field

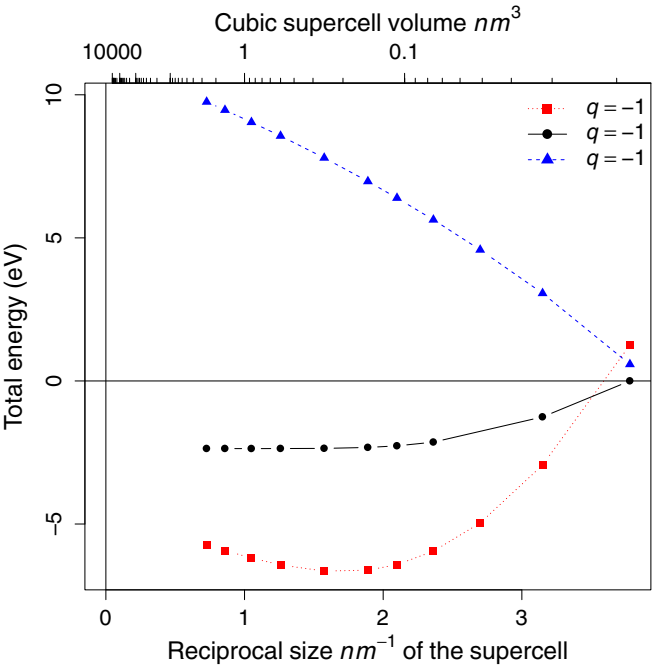

FIG. 2. (Color online) Supercell-size dependence of the total energy of an oxygen atom in vacuum. Three charge states q are compared.

$u(r)$ of a monopole source [100,101] varies with the distance r in the same way as the electrostatic field:

$$u(r) = \frac{\Omega}{4\pi k r^2}, \tag{7}$$

and the energy density of the strain field is proportional to $u(r)$ and always positive, as in the case of electrostatic energy density [102]. The defect volume Ω is thus an elastic equivalent of the electrostatic charge q. Employing this similarity, one can immediately see that although spherical inclusions embedded in an isotropic elastic medium repel one another, the elastic monopole term is attractive as well. The elastic analog of the neutralizing background used in supercell calculations of charged objects is the constraint of constant cell dimensions: the supercell translation vectors are not relaxed when the defect is introduced, but they are kept equal to that of the perfect crystal. In other words, the stress field induced at the cell boundary by the point defect is ignored. This stress may be computed, but it has no physical meaning.

The monopole term dominates asymptotically. At finite distances, the cell-size dependence may be influenced by the r^{-3} term and by higher-order multipole interaction and also by other effects. For example, wave functions of localized states decay exponentially with distance; hence, the effect of their overlap might be approximated by an exponential term [78]. This is visible in Fig. 2 for small cells in the behavior of the neutral O atom and of the negatively charged O atom (on which the added electron occupies an extended wave function).

There are also at least two other sources of size dependence of the total energy. First, as mentioned above, small size-dependent inaccuracies in the description of a single atomic bond in the crystal add up in large cells, so that there is a size-dependent k-point sampling error [78]. Since the quality of any k-point sampling scheme tends asymptotically to perfectness, this error must vanish at infinity, yet it is not immediately obvious that the error per atom decreases with the cell volume much faster than the number of atoms does. For example, it was reported that Γ sampling in sc cells introduces a r^{-2} term [78].

Second, charged cells are neutralized by setting the divergent zero-order Fourier term to zero, rather than adding a homogeneous background charge [103]. The side effect is that the average electrostatic potential $\overline{\Phi}$ that determines the reference energy for E_F becomes ill defined, so that the total energy of a defect in the charge state q is shifted by $q\Delta\overline{\Phi}$ [77]. The value of $\Delta\overline{\Phi}$ may be estimated by comparing the electrostatic potential far away from the defect to that of the uncharged bulk [77,78,104]. This estimate may be, however, quite inaccurate [74]. Fortunately, one can argue that $\Delta\overline{\Phi}$ decays proportionally to the number of atoms [75,78], i.e., as r^{-3}, so that this effect may be treated as a contributor to the coefficient B in Eq. (6).

All this makes a potentially complex picture. Yet since the number of manageable cell sizes is, for Si calculations, small (fcc periodicity may, in practice, vary from 4 up to 9, and in many cases we varied it only up to 7), the fitting function must be simple. We have therefore extrapolated the formation energies E_{form} to infinity using the formula

$$E_{\text{form}}(r) = E_\infty - |A| r^{-1} + f_{\text{core}}(r), \tag{8}$$

where $f_{\text{core}}(r)$ describes the contributions beyond the monopole-monopole interaction, depends on one or, at most, two parameters, and vanishes with $r \to \infty$ faster than r^{-1}. The following forms of $f_{\text{core}}(r)$ have been tried:

$$f_{\text{ri}}(r) = B r^{-i}, i = 2, \ldots, 11\text{: models r}i, \tag{9}$$

$$f_{\text{p}}(r) = B x^{-|C|}\text{: model p}, \tag{10}$$

$$f_{\text{ei}}(r) = B \exp(-|C| r^i), i = 1,2\text{: models e}i. \tag{11}$$

The fit parameters are E_∞, A, B, and C. We used the PORT algorithm [105] as implemented in the NLS method of the R language [106] and adapted the Monte Carlo strategy.

It often turns out that the dependence of the long-range interaction parameter A [Eq. (8)] on the charge q of the defect can, with good accuracy, be approximated by

$$A(q) = A_0 + A_1 q + A_2 q^2. \tag{12}$$

This is not surprising because the electrostatic contribution to A has the form of $A^{\text{ele}} = A_2^{\text{ele}} q^2$, and the elastic contribution has the form of $A^{\text{ela}} = A_2^{\text{ela}} \Omega(q)^2$. One can write $\Omega(q) = \Omega_0 + \Sigma_i^\infty \omega_i q^i$; the terms Ω_0 and ω_1 are responsible for A_0 and A_1 being nonzero, and if the product $\Omega_0\omega_2$ is negative, it may overweight the positive A_2^{ele}, causing the net value of A_2 to become negative. The $A(q)$ dependence can be monitored in order to estimate the upper limit for the contribution of electrostatics to the supercell-size convergence. (For an example, see Appendix B.)

Finally, in order to visualize and, as far as possible, to measure the degree of structural relaxation of the defect core, we also followed the cell-size dependence of selected inter-atomic distances. The converged distances d_∞ were approximated by fitting two arbitrary test functions:

$$f_{\text{bond}}(r) = d_\infty[1 + \exp(-A r^B)], \tag{13}$$

$$g_{\text{bond}}(r) = d_\infty + A r^B. \tag{14}$$

III. RESULTS FOR FINITE SUPERCELLS

Table II lists the formation energies of the vacancy and its clusters in stable and in (some) nearly stable structures. The column "class" describes the cluster type. For V_1 and V_2 it is the point group (pg) symmetry. The class "hex" consists of vacancies occupying the lattice sites of a hexagonal Si ring. The pivotal cluster is here the V_6 cluster, with the whole ring removed [Figs. 3(a) and 3(b)]. Part of the hexagonal ring (PHR) hex substructures are made by removing $3 \leqslant n \leqslant 5$ atoms from this ring. In fourfold coordinated (FFC) substructures, pairs of atoms are removed, and single atoms are reinserted in between. For example, $V_{5,\text{FFC}}$ is built by removing the atom pair (1, 2) and reinserting a Si atom between the vacant sites. The pivotal cluster in the "void" class is the V_{10} cluster that forms a perfect octahedral void in the Si bulk [Figs. 3(c) and 3(d)].

The energies in Table II are computed at Fermi level $E_F = 0$ [GGA VBM, $dE_F = 0$, Eq. (4)]. E_{form} at arbitrary E_F can be obtained by adding to $E_{\text{form}}(E_F{=}\,0)$ the term qE_F, where q is the charge state of the cluster [Eq. (1)].

TABLE II. E_{form} of stable and nearly stable V_n, calculated for E_{F} at the GGA VBM. EPR-active charge states are marked by the up arrow (↑). For the dilute limit, see Table VIII.

Defect	Class	Charge	4 128	5 250	6 432	7 686	8 1024	9 1458
V_1	T_d	2	2.51	3.03	3.28	3.41	3.50	3.56
V_1	D_{2d}	↑ 1	2.84	3.24	3.43	3.53	3.59	3.62
V_1	D_{2d}	0	3.19	3.42	3.52	3.59	3.62	3.64
V_1	C_{2v}	↑ −1	3.66	3.84	3.91	3.97		
V_1	D_{2d}	↑ −1	3.68	3.86	3.92	3.98	4.01	4.02
V_1	D_{3d}	↑ −1		4.01	4.04	4.07		
V_1	D_{3d}	−2	4.28	4.30	4.30	4.34	4.37	
V_1	D_{2d}	↑ −2	4.21	4.36	4.41	4.47	4.51	
V_2	C_{2h}	↑ 1	4.29	4.89	5.15	5.28	5.35	5.40
V_2	C_{2h}	↑ 0	4.26	4.92	5.17	5.30	5.37	5.41
V_2	C_{2h}	↑ −1	4.34	5.03	5.33	5.49	5.58	5.63
V_2	D_{3d}	−2	4.48	5.24	5.58	5.78	5.90	5.96
V_3	PHR	2	5.82	6.50	6.80	6.93	7.02	
V_3	PHR	↑ 1	5.72	6.57	6.84	6.95	7.02	
V_3	PHR	↑ 0	5.71	6.64	6.88	6.99	7.06	
V_3	PHR	↑ −1	5.81	6.79	7.15	7.31	7.35	7.38
V_3	PHR	−2	5.98	6.95	7.43	7.58	7.66	7.70
V_4	FFC	1	7.18	7.49	7.58	7.61		
V_4	FFC	0	6.76	7.27	7.46	7.54		
V_4	FFC	↑ −1	6.48	7.34	7.68	7.84		
V_4	FFC	−2	6.24	7.42	7.92	8.19		
V_5	FFC	1	7.98	8.55	8.75	8.82		
V_5	FFC	0	7.44	8.28	8.59	8.72		
V_5	FFC	↑ −1	7.29	8.43	8.86	9.06		
V_5	FFC	↑ −2	7.16	8.56	9.15	9.45		
V_6	hex	1	8.80	9.48	9.68	9.74	9.77	
V_6	hex	0	8.14	9.15	9.49	9.62	9.70	
V_6	hex	↑ −1	7.94	9.25	9.73	9.95	10.07	
V_7	void	0	9.86	11.20	11.68	11.88		
V_7	void	↑ −1	9.55	11.13	11.70	11.96		
V_7	void	−2	9.27	11.11	11.83	12.18		
V_7	void	↑ −3	9.12	11.32	12.20	12.65		
V_8	void	↑ 0	10.67	12.07	12.71	12.93		
V_8	void	↑ −1	10.10	11.77	12.48	12.78	12.93	
V_8	void	↑ −2	9.61	11.60	12.41	12.82	13.02	
V_8	void	↑ −3	9.07	11.28	12.25	12.76	13.05	
V_8	void	−4	8.52	11.17	12.31	12.93	13.29	
V_8	void	↑ −5	8.32	11.32	12.69	13.44	13.90	
V_9	void	↑ 0	11.35	12.72	13.25	13.41	13.50	
V_9	void	↑ −1	10.70	12.39	12.98	13.24		
V_9	void	↑ −2	10.11	12.10	12.89	13.26		
V_9	void	↑ −3	9.48	11.82	12.78	13.28		
V_9	void	−4	8.82	11.51	12.68	13.32		
V_9	void	↑ −5	8.42	11.66	13.09	13.86		
V_{10}	void	↑ 0	11.98	13.13	13.50	13.63	13.69	
V_{10}	void	↑ −1	11.31	12.86	13.37	13.61	13.73	
V_{10}	void	↑ −2	10.63	12.51	13.24	13.76	13.78	
V_{10}	void	↑ −3	9.91	12.16	13.10	13.58	13.85	
V_{10}	void	−4	9.17	11.80	12.95	13.57	13.94	
V_{10}	void	−5	8.68	11.97	13.39	14.13	14.58	
V_{11}	void	↑ 0	13.16	14.48	14.92	15.25	15.17	15.20
V_{11}	void	↑ −1	12.31	14.10	14.77	15.06	15.21	
V_{11}	void	−2	11.44	13.72	14.61	15.04	15.27	
V_{11}	void	↑ −3	10.93	13.51	14.51	15.02	15.30	
V_{11}	void	−4	9.75	13.48	14.61	15.21	15.56	
V_{11}	void	↑ −5	9.31	13.59	14.98	15.74	16.20	

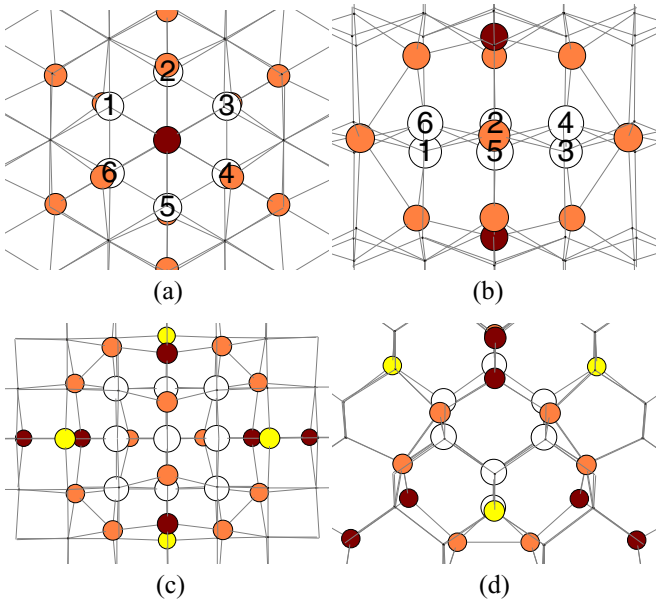

FIG. 3. (Color online) Pivotal V_n clusters. Vacant sites are indicated as white circles. Dimerized atoms are yellow (light gray), atoms with broken bonds are orange (medium gray), and atoms with strongly strained bonds are brown (dark gray). V_6, the generator of the "hex" (FFC, PHR) class, viewed (a) along (111) and (b) along $(1\bar{2}1)$. V_{10}, the generator of the "void" class, viewed (c) along (001) and (d) along (110).

A. Single vacancy

V_1 orbitals are built by hybridization of the four sp^3 dangling bonds around the vacant site. They can hold up to eight electrons and are occupied by four in uncharged V_1^0, so the charge state q might thus vary between +4 and −4. But adding or removing electrons leads to electrostatic repulsion. The orbital energy increases, and the charge transition level then moves up ($q < 0$) or down ($q > 0$); see Eq. (3). When $|q|$ increases too much, the level leaves the gap, and the excess charge is released. Ultimately, the charge of V_1 can vary between +2 and −2 [2], whereby V_1^+ is unstable against $2V_1^+ \rightarrow V_1^0 + V_1^{2+}$ (Anderson negative-U system) [2,4,14,15,18,21,28,30,37]. The symmetry of V_1^+ is D_{2d} (two identical dimers), and that of V_1^- is C_{2v} (a shorter and a longer dimer), as determined by EPR [2,36]. The other charge states are not EPR active, but from the tight-binding picture [2] and from DFT [28,30] it follows that V_1^{2+} is T_d (no dimerization) and V_1^0 is D_{2d}. According to DFT [21,28,30], V_1^{2-} is special in that an atom moves towards the other three to make six identical bonds in a quasihexagonal "split-vacancy" configuration of D_{3d} symmetry [3].

Our results agree with these facts. We also confirm the existence of several metastable structures with the charge other than +2 that have nearly the same energy [24,25,28,29]. The length of the dimer in V_1^0 decreases with the cell size and saturates at about 0.31 nm. The convergence is achieved faster in the sc superlattice [Fig. 4(a)]. The structure is converged with respect to the basis set [Fig. 4(a)]. Apart from the 128-site cell, atoms hardly move when a denser k-point set is used (2^3 MP grid with a half-grid step offset) [107]; the energy gain from the structure relaxed at Γ to that relaxed at 2^3 MP goes

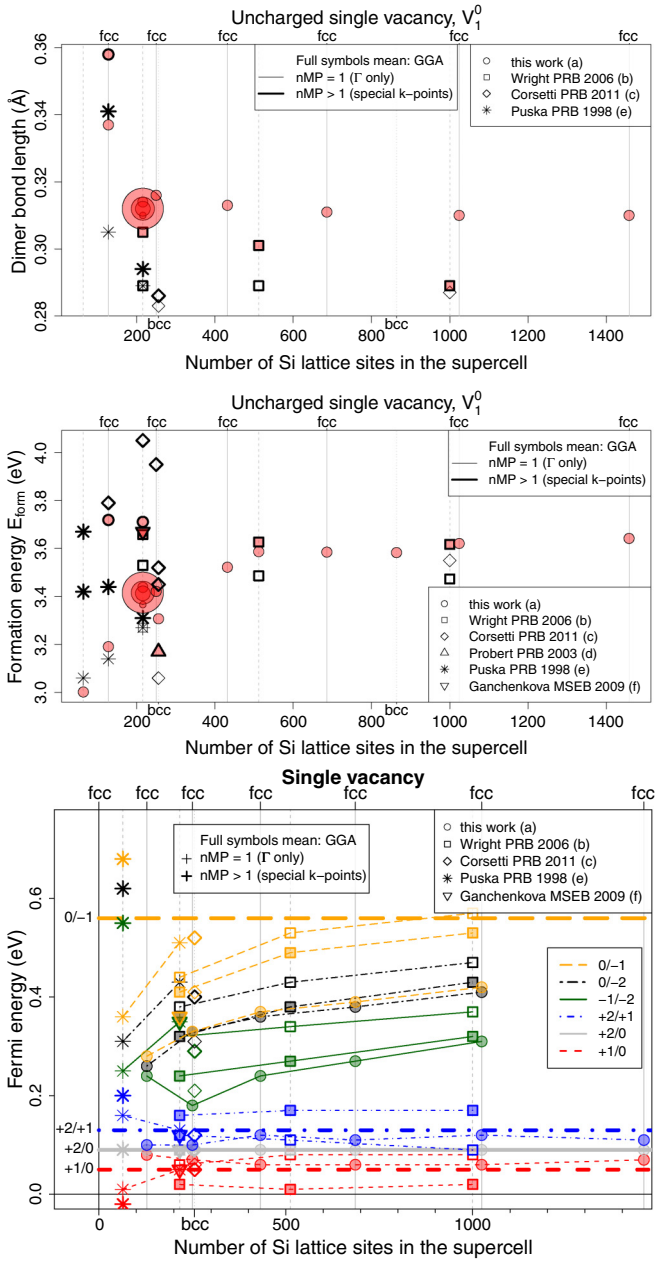

FIG. 4. (Color online) Dependence of various vacancy parameters on the cell size, the basis set (8–20 Ry for sc cell with 216 sites; circle size reflects the basis set size), Bravis superlattice (sc, fcc, bcc), and BZ sampling (Γ, 2^3 MP), and comparison to literature. (a) V_1^0 dimer length. (b) $E_{\rm form}(V_1^0)$. (c) Gap levels, aligned as in Table III. References: b, [28]; c, [30]; d, [24]; e, [21]; f, [29].

down from 90 meV for the 128-site cell through 3 meV in the 250-site cell to 0 meV in the 1458-site cell. In most cases, the dimers are longer than those reported so far [Fig. 4(a)]. This is not due to metastability and may be caused by differences in the bulk elastic properties for various pseudopotentials. The effect is too small to be reflected in formation energies [Fig. 4(b)].

Level energies obtained from Eq. (3) agree with those computed by other authors (Fig. 4 and Table III). For better comparison, Fermi level correction $dE_{\rm F}$ has been applied to align all DFT (2+/0) levels at the experimental value. When computed for our GGA levels, $dE_{\rm F}$ decreases with the cell size from 0.48 eV for the smallest 128-site cell and 0.32 eV for the 250-site cell down to 0.16 and 0.15 eV for the 1024- and 1458-site cells, reflecting the decrease of the raw (2+/0) energy with increasing cell size. Anticipating further discussion, we note that the most appropriate correction (with the smallest sum of squared residuals) for V_1 and V_3 defect levels at the dilute limit is $dE_{\rm F} = 0.19$ eV, which is only marginally higher than the 0.14–0.15 eV expected from this behavior.

TABLE III. Vacancy gap levels at finite cells, compared to literature (bold font: this work). Unstable levels are in italics. DFT levels are aligned such that all (2+/0) levels match the experiment. Exchange-correlation type, cell size, k-point sampling, and reference are indicated. An asterisk (∗) means that acceptor levels (extrapolated [108] from SiGe) are recalculated to the position above the VBM, assuming a gap of 1.12 eV [84].

Vacancy	+/0	‡/0	‡/+	−/=	0/=	0/−
Experiment	*0.05* [5]	0.09 [5]	*0.13*[5]	0.85*[108]		0.56*[108]
GGA^{Γ}_{1024}	***0.06***	**0.09**	***0.12***	***0.31***	**0.41**	***0.42***
GGA^{MP}_{1000} [28]	*0.02*	0.09	*0.17*	*0.32*	0.43	*0.53*
LDA^{MP}_{1000} [28]	*0.08*	0.09	*0.09*	*0.37*	0.47	*0.57*
GGA^{Γ}_{250}	***0.07***	**0.09**	***0.10***	*0.18*	**0.33**	***0.33***
LDA^{Γ}_{256} [30]	*0.06*	0.09	*0.12*	*0.21*	0.31	*0.41*
LDA^{MP}_{256} [30]	*0.05*	0.09	*0.12*	*0.29*	0.40	*0.52*
GGA^{MP}_{216} [29]	*0.05*	0.09	*0.12*	*0.35*	0.36	*0.36*
GGA^{MP}_{216} [28]	*0.02*	0.09	*0.16*	*0.24*	0.32	*0.41*
LDA^{MP}_{216} [28]	*0.06*	0.09	*0.12*	*0.32*	0.38	*0.44*
LDA^{Γ}_{216} [21]	*0.05*	0.09	*0.13*	*0.35*	0.43	*0.51*
LDA^{Γ}_{64} [21]	*0.01*	0.09	*0.16*	*0.25*	0.31	*0.36*
LDA^{MP}_{64} [21]	*−0.02*	0.09	*0.20*	*0.55*	0.62	*0.68*

From Table III it follows that the energy order and distance of DFT donor levels compare well to experiment. Even 64 sites with Γ point work (the 3^3 Monkhorst-Pack grid yields worse results for such a small cell) [21]. The donor states in the biggest cells are probably affected by the artificial interaction with the GGA valence band. They are preferably ignored in the extrapolation to the dilute limit (Fig. 5) and the power n in the $(q_1/q_2) + AN^{-1} + N^{-n}$ dependence should be at least 4 (model r4 or higher); otherwise, the extrapolation reverses the sign of Anderson U. Figure 5 adapts r4 for the donor states (with $9 \times 9 \times 9$ data ignored) and r3 [on average, the best for $E_{\rm form}(V_1^q)$] for the acceptor states.

DLTS levels of the acceptor states cannot be measured directly [108]. Extrapolation [108] from SiGe [109] may be in conflict with the negative U predicted by all DFT calculations for the acceptor states. The experimental assignment of charge states to the V_1 acceptors is tentative [110], but if it is correct, the reason for the discrepancy may be that local strain around Ge atoms (to which V_1 is attracted) [108] affects the order of the gap states; indeed, DFT sees that strain influences the order of states, at least for V_2 [54]. When the D_{3d} states are omitted [111], the acceptor levels compare favorably to DLTS extrapolation from SiGe, particularly in the dilute limit (Fig. 5).

B. Divacancy

When two neighboring atoms are removed from the bulk, a nearest-neighbor divacancy (nnV_2) is formed. In contrast to

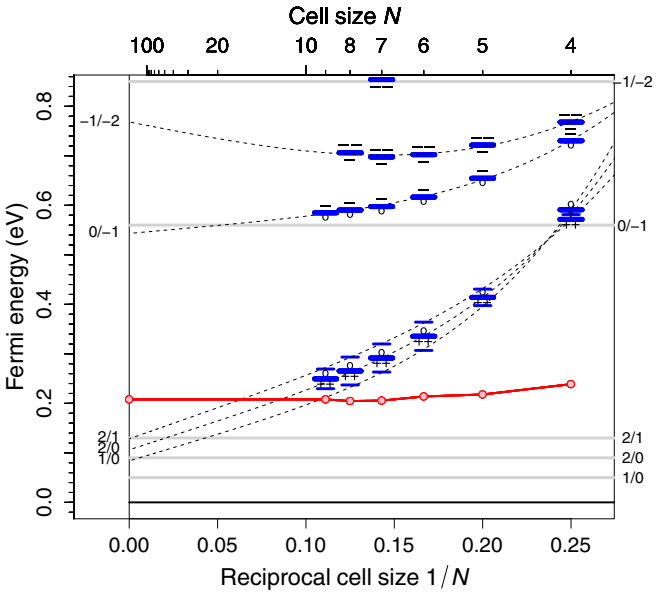

FIG. 5. (Color online) V_1 levels without the D_{3d} state; $dE_F = 0.19$ eV. Long dashes indicate stable levels. The interconnected points mark the GGA VBM, and gray lines show the experiment [5,108].

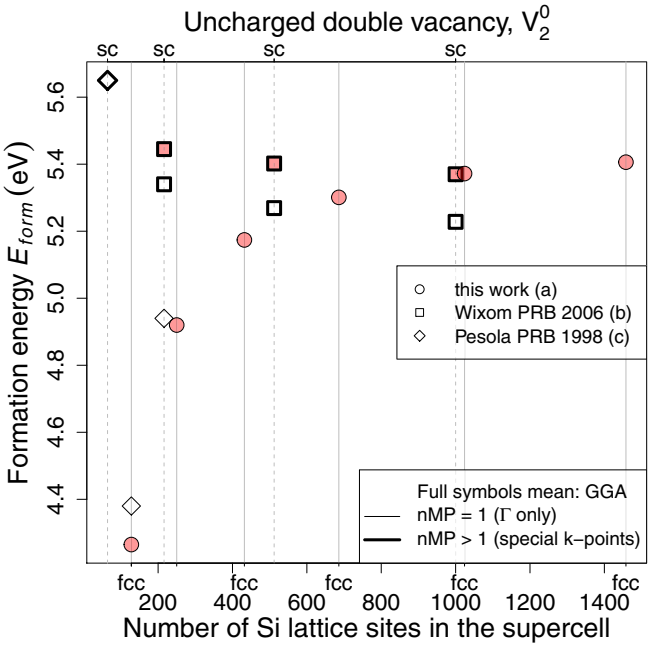

FIG. 6. (Color online) E_{form} of uncharged V_2, compared to Refs. [53] (labeled b) and [51] (labeled c).

V_1, not all broken bonds of the neighboring atoms of nnV_2 can be recombined into dimer bonds. Only two second-neighbor (2nn) dimers can be built. Dimerization of the remaining atoms leads to third-neighbor (3nn) bonds, which introduces so much strain that the structure spontaneously reverts to a geometry with all 3nn bonds opened. This happens in small as well as in big cells. The nnV_2 geometry is also spontaneously (i.e., without energy barrier) acquired when the removed atoms were the second, not the first, nearest neighbors.

V_2 bonding depends on the charge state [1,40,43,45,47–54]. This is our observation as well. We also confirm that for charge states other than V_2^{2-} there are various metastable bonding distortions of Jahn-Teller origin that differ in energy so little that they may be difficult to differentiate. For example, in all cells studied here the energy difference ΔE between the resonant and nonresonant bonding for V_2^1 comes out smaller than 10 meV. For V_2^0, ΔE may reach 40 meV, but it changes sign as the cell increases in size. Since in our current calculations the inaccuracy due to finite basis set is about 20 meV per vacancy [112], we do not distinguish here between these variations. For the purpose of this study, we focus on the nonresonant distortion mode with one shorter and two longer bonds [40,47].

Figure 6 compares the size dependence of the uncharged divacancy obtained here for fcc cells with Γ sampling to published data [51,53]. It is apparent that there is a common asymptotic behavior and that from about 1000 atoms on all GGA data sets would follow the same line. The structure of V_2^0 is converged with respect to BZ sampling: relaxation with the 2^3 set from the Γ-optimized structure reduces the energy by 50, 17, 3 and 0 meV for the 128-, 250-, 432-, and 683-site cells, respectively.

Figure 7 and Table IV collect the results on the divacancy gap levels. The positions of acceptor states calculated in this work compare well to the DLTS data [113]. The distance between the two acceptor levels hardly varies with the cell size and is close to its experimental value [Fig. 7(a), Table IV]. The absolute positions of the acceptor levels in the gap coincide with experiment when the valence-band edge correction of $dE_F = 0.36$ eV is applied and the levels are extrapolated to cells with $N = \infty$ atoms using the r4 model that, on average, performs the best for $E_{form}V_2^q$ [Fig. 7(a) and Sec. IV]. The correction of $dE_F = 0.36$ eV places the GGA and true conduction-band (CB) edges at the same energy, indicating that the CB contribution to the V_2^q states for $q \geqslant 0$ is predominantly correct in our GGA calculations. But the donor level appears to be poorly described due to artificial admixture of valence states; this is a GGA artifact.

Our GGA calculations at Γ reproduce the correct ordering of the acceptor gap levels even in small fcc cells (128 sites). This is in contrast to the GGA calculations with sc cells and a formally converged k-point sampling [53], which reverse the order of the acceptor states even in the large sc cell of 1000 sites [Fig. 7(b), Table IV]. The experimental order is recovered only after extrapolation to the dilute limit [53]. This may be a combined effect of BZ sampling and the Bravis superlattice because the LDA calculations done with fcc cells and away from the Γ point reproduce the experimental order [51], while changing GGA to LDA in sc cells with a non-Γ sampling is no remedy [53].

In spite of that, the GGA formation energy of V_2^0 is already practically the same when computed for the 1000-site sc cell with a converged k-point set and for the 1024 fcc cell with only Γ (Fig. 6). This illustrates the fact that a simple verification that the defect in a single charge state converges at certain conditions does not prove that these conditions will yield good gap levels as well.

C. Trivacancy

Trivacancy (V_3) is the third of the Si vacancy defects, for which the charge transition levels have been identified

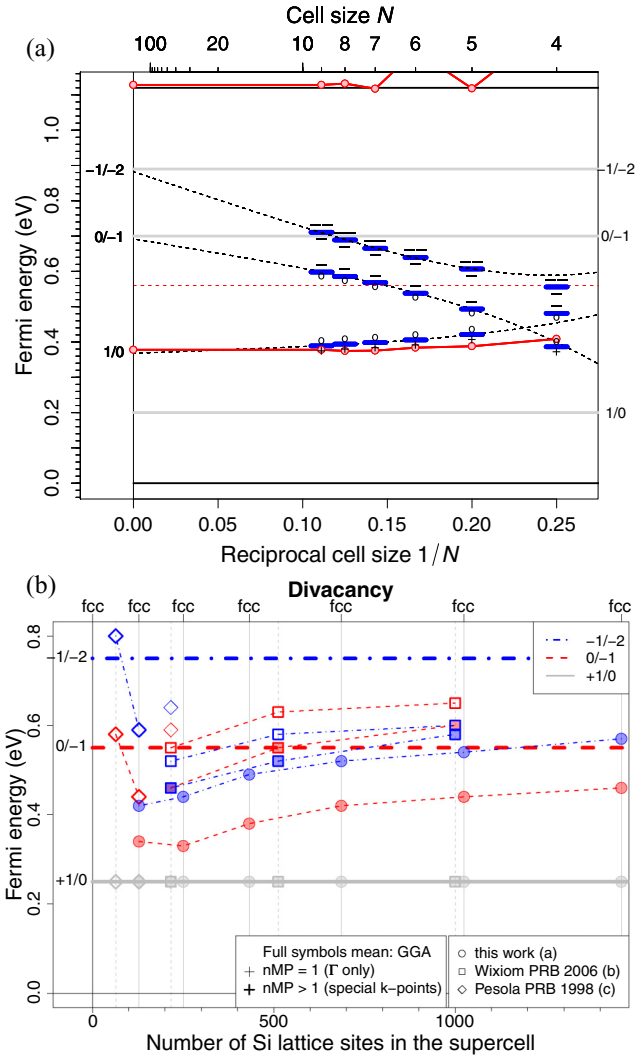

FIG. 7. (Color online) Convergence of divacancy gap levels. Gray lines mark the DLTS energies [45]. (a) GGA levels obtained in this work and adjusted by $dE_F = 0.36$ eV. Fits are done with the r4 formula [Eq. (9)] directly to levels, ignoring the $4 \times 4 \times 4$ data. The interconnected points are the GGA band edges. (b) DFT levels from Table IV. References: b, [53]; c, [51].

by DLTS [55–58]. V_3 may exist in two different atomic configurations, both of which are built on the basis of the hexagonal ring [Fig. 3(a)]. The ground state of V_3^0 is known as the FFC structure, produced by removing each second atom from the ring. The other stable states of trivacancy are V_3^- and V_3^{2-}. Their ground state is known as the PHR structure; it is produced by removing three subsequent atoms from the ring. DFT consistently returns $E_{\text{form}}V^0_{3,\text{PHR}} > E_{\text{form}}V^0_{3,\text{FFC}}$ [Fig. 8(a)].

The core of the FFC type of V_3 [Fig. 8(b)] converges with supercell size more smoothly than the core of the PHR V_3 does [Fig. 8(c); the small jump between the two biggest cells that is seen in Fig. 8(b) is already within the numerical accuracy range]. Namely, the bonding pattern (and, with it, the symmetry) of the PHR cluster in its ground state varies with the cell size and stabilizes only in $7 \times 7 \times 7$ and larger cells, whereby the interatomic distances appear to practically converge as well. These variations contribute to the inaccuracy in the determination of the converged energies and charge transition levels. For this reason, V_3 seems to be a difficult case.

It may be surprising that there is a substantial difference between the transition levels obtained in Ref. [58] and in this work using $dE_F = 0$ (Table V, the values set in italics). This difference exemplifies the effect of BZ sampling. Both calculations use a single k point, but while in this work it is the Γ point (BZ center), in Ref. [58] it is the L point (BZ boundary). Taking the L point widens the gap between the available Bloch states. The highest occupied state of the sampled perfect crystal states is now away from the DFT VBM at Γ, and the lowest empty state of the sampled perfect crystal is farther away from the DFT conduction band maximum (CBM) at the Δ line than when Γ sampling is used. First, this means that the reference point [the "hole" DFT "VBM", Eq. (4)] is different in both calculations. Second, the wider separation between the donor and acceptor states of V_3 obtained at L indicates that the admixture of band states to the acceptor states is significant. Still, extrapolation of Γ results to infinity with $dE_F = 0.17$ (which is nearly the same as that for V_1) fully reproduces the DLTS levels (Table V). The better agreement of the L-point data with experiment may be due to partial cancellation of errors and can be viewed as an outcome of incomplete convergence. But precisely due to this cancellation, the 512-atom sc cell with L sampling [58] may be good for DFT calculations for point defects in Si.

It follows that the contribution of the DFT band gap problem to the inaccuracy in the computed V_n levels cannot be neglected. In addition to the uncertainty in the reference level, there is an uncertainty due to the admixture of band states that affects the E_{form} directly [79,115].

TABLE IV. Gap states of the divacancy at finite cells (bold font) compared to literature data. DFT levels are aligned such that all (+/0) levels match the DLTS data. Exchange-correlation type, cell size, k-point sampling (Γ or a denser MP grid), and literature source are indicated. By the data from Refs. [40], [43], and [45], an asterisk (∗) means that acceptor levels are recalculated to the distance from the VBM assuming a gap of 1.12 eV [84].

Divacancy	+/0	0/−	−/=
EPR/PAS	0.25 [40]	0.55 [46]	0.75* [40,46]
DLTS	0.20 [43,45]	0.70* [43,45]	0.89* [43,45]
$\text{GGA}^{\Gamma}_{1458}$	**0.20**	**0.41**	**0.52**
$\text{GGA}^{\Gamma}_{1024}$	**0.20**	**0.39**	**0.49**
$\text{GGA}^{\text{MP}}_{1000}$ [53]	0.20	0.55	0.53
$\text{LDA}^{\text{MP}}_{1000}$ [53]	0.20	0.60	0.55
$\text{GGA}^{\Gamma}_{686}$	**0.20**	**0.37**	**0.47**
$\text{GGA}^{\text{MP}}_{512}$ [53]	0.20	0.55	0.52
$\text{LDA}^{\text{MP}}_{512}$ [53]	0.20	0.58	0.53
$\text{GGA}^{\Gamma}_{432}$	**0.20**	**0.33**	**0.44**
$\text{GGA}^{\Gamma}_{250}$	**0.20**	**0.28**	**0.39**
$\text{GGA}^{\text{MP}}_{216}$ [53]	0.20	0.46	0.46
$\text{LDA}^{\text{MP}}_{216}$ [53]	0.20	0.50	0.47
$\text{LDA}^{\Gamma}_{216}$ [51]	0.20	0.54	0.59
$\text{GGA}^{\Gamma}_{128}$	**0.20**	**0.29**	**0.39**
$\text{LDA}^{\text{MP}}_{128}$ [51]	0.20	0.39	0.54
$\text{LDA}^{\text{MP}}_{64}$ [51]	0.20	0.53	0.75

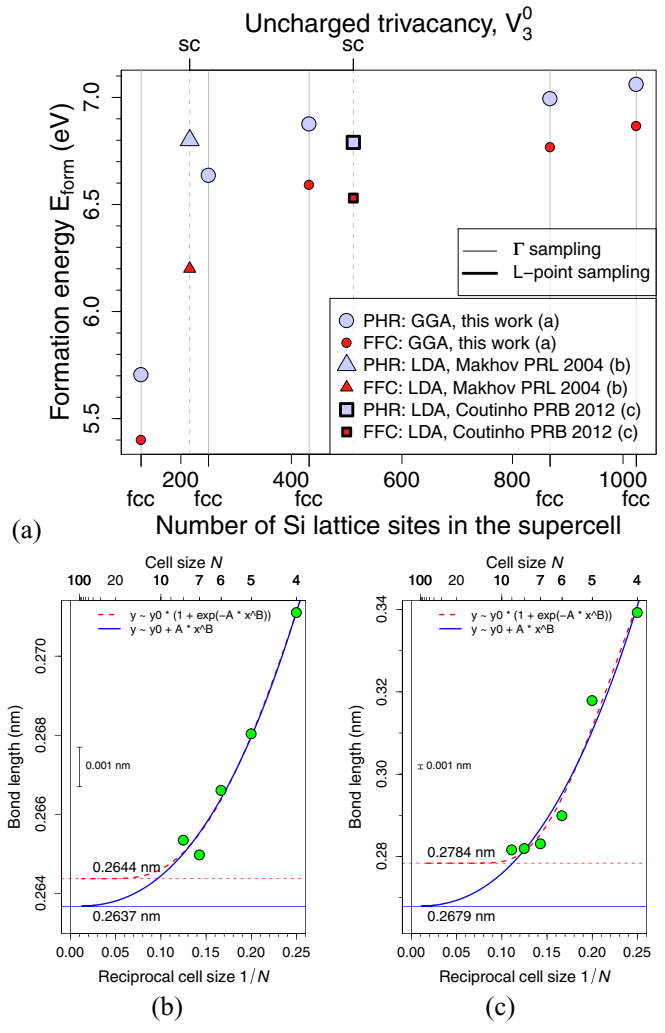

FIG. 8. (Color online) Cell-size convergence of V_3 calculations. (a) $E_{\rm form}$ of uncharged V_3, compared to Refs. [59] (labeled b) and [58] (labeled c). Dimer bond length (b) of FFC V_3^0 (this work) and (c) of PHR V_3^- (this work).

TABLE V. Charge transition levels of Si trivacancy compared to experimental data (DLTS, Refs. [56,57]) and to the LDA results obtained in Ref. [58] for sc cells and BZ sampling at the L point [114]; N is the cell size. DFT levels in roman are aligned by $E_{\rm F}$ shift $dE_{\rm F}$ to match the (1/0) DLTS energy. The levels in italics correspond to $dE_{\rm F} = 0$.

Trivacancy	++/+ P/P	+/0 P/P	0/− P/P	−/= P/P
DLTS	0.106	0.193	0.662[a]	0.761[a]
GGA, $N = \infty$	0.13	0.19	0.68	0.75
GGA, $N = 1024$	0.16	0.19	0.45	0.46
GGA, $N = 686$	0.18	0.19	0.47	0.42
LDA, $N = 512$	0.05	0.19	0.80	0.92
GGA, $N = 686$	*0.02*	*0.04*	*0.32*	*0.27*
LDA, $N = 512$	*0.15*	*0.29*	*0.90*	*1.02*

[a]DLTS levels are recalculated from the distance from the CBM to the distance from the VBM, assuming a gap of 1.12 eV.

TABLE VI. $E_{\rm form}$ of $V^0_{4,\rm aFFC}$ and $V^0_{4,\rm FFC}$ (both spin unpolarized) and of metastable tetra- and pentavacancy configurations in various charge states. $E_{\rm F}$ is at GGA VBM. EPR-active charge states are indicated by the up arrows (↑). aFFC: asymmetric FFC.

Defect	Type	Charge	Spin polarized allowed	4 128	5 250	6 432	7 686
V_4	aFFC	0	N			7.617	
V_4	FFC	0	N			7.541	
V_4	110	↑ 1	Y	7.325	8.015	8.304	8.439
V_4	110	↑ 0	Y	7.140	8.126	8.366	8.701
V_4	110	↑ −1	Y	7.076	8.254	8.809	9.039
V_4	110	−2	Y	6.970	8.467	8.994	9.264
V_4	void	↑ 1	Y	7.461	8.483	8.938	9.145
V_4	void	↑ 0	Y	7.359	8.453	8.879	9.080
V_4	void	↑ −1	Y	7.420	8.634	9.055	9.269
V_4	void	↑ −2	Y	7.603	8.846	9.309	9.558
V_5	void	↑ 1	Y	8.856	10.124	10.576	
V_5	void	↑ 0	Y	8.610	10.016	10.476	
V_5	void	↑ −1	Y	8.548	10.007	10.579	

D. $V_4 - V_{11}$ clusters

Vacancies in complexes containing up to seven single defects tend to arrange themselves in rings built on the basis of the quasihexagonal (111) rings in the diamond structure [59,60,62–65]. This can be visualized on the basis of the ringlike V_6 complex. In V_6, all atoms from a single ring are removed, and their neighbors are rebonded into six dimers [Figs. 3(a) and 3(b)]. In $V_{5,\rm FFC}$, an atom is reinserted between sites 1 and 2 [Figs. 3(a) and 3(b)]. In V_4, one more atom is returned and is placed in a symmetric [between sites 4 and 5, Figs. 3(a) and 3(b)] or asymmetric [between sites 3 and 4, Figs. 3(a) and 3(b)] manner. Table VI lists the computed $E_{\rm form}$, and Fig. 9 compares the $E_{\rm form}(V_n^0)$ with the literature data. Metastable structures are covered as well.

For V_4, the asymmetric hex ring ($V_{4,\rm aFFC}$) [59,64], the ⟨110⟩-planar chain ($V_{4,110}$) [55,59,64] assigned to the P3 spectrum in the EPR signal of neutron-irradiated Si [55], and the void-type ⟨111⟩-planar chain ($V_{4,\rm void}$) [55] assigned to the A3 spectrum [55] belong to the metastable group. In accord with Ref. [59], we find that the symmetric structure is slightly (by about 0.15 eV) more stable than the asymmetric one. For V_5, the structure designated "void" is the nonplanar pentavacancy assigned to the P-1 EPR spectrum from neutron-irradiated Si [68].

Clusters from V_7 to V_{11} (which is the largest cluster considered in this work) can be built on the basis of two hexagonal rings sharing one site [see Fig. 9(d), where vacant sites in the second ring are shown as dark circles]. Two such rings form the perfect $V_{11,\rm hex}$. Again, PHR and FFC substructures are possible for intermediate n. But beyond six vacancies, the ring-based complexes are metastable (Table VII) and the stable complexes form on the basis of the octahedral void [Figs. 3(c), 3(d), and 9(e)].

In variation to the hex PHR structures, the presence of threefold-coordinated atoms [Figs. 3(c), 3(d), and 10(a)] is typical for the void class. They hold electrons easily, so that the donor levels of these clusters are hidden in the valence

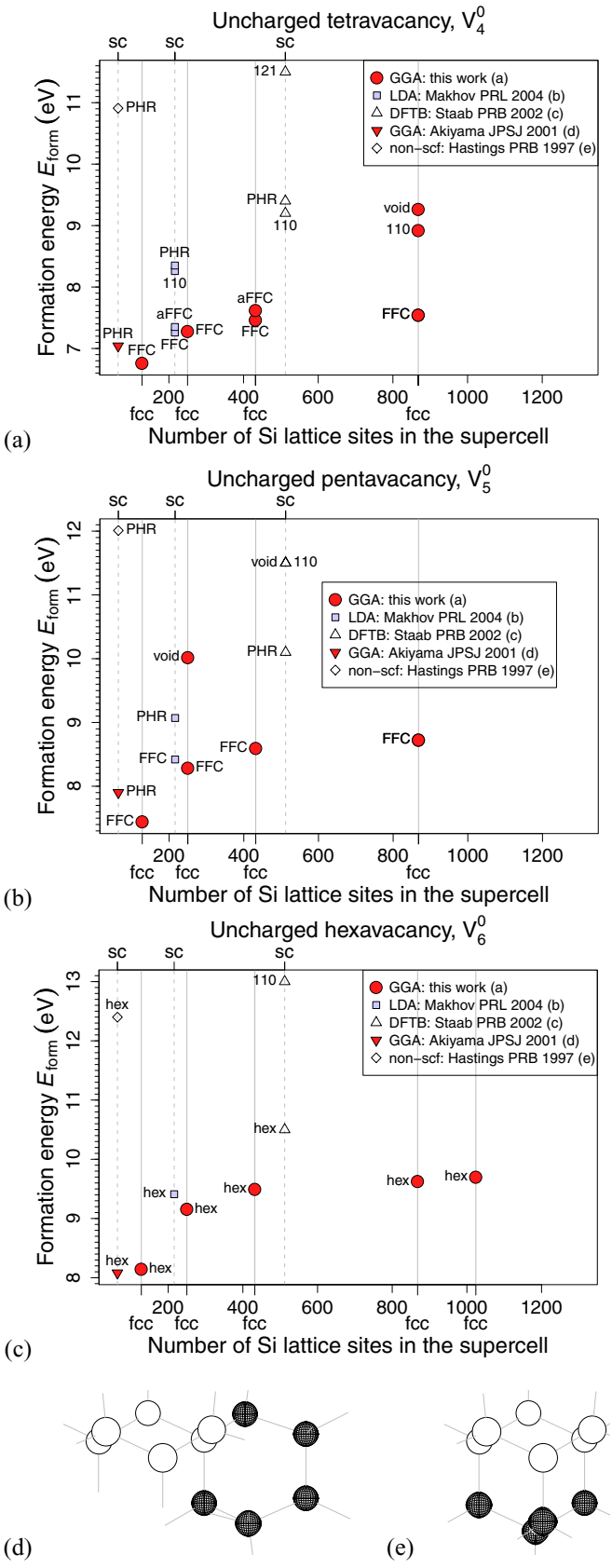

FIG. 9. (Color online) (a)–(c) Cell-size convergence of $E_{\text{form}} V_n^0$ for $4 \leqslant n \leqslant 6$; data from b, [59]; c, [64]; d, [66]; e, [65]. (a) V_4^0, (b) V_5^0, and (c) V_6^0. (d) Two interwoven vacant hexagonal rings: perfect $V_{11,\text{hex}}$. (e) Additional vacant sites under the ring: perfect $V_{10,\text{void}}$.

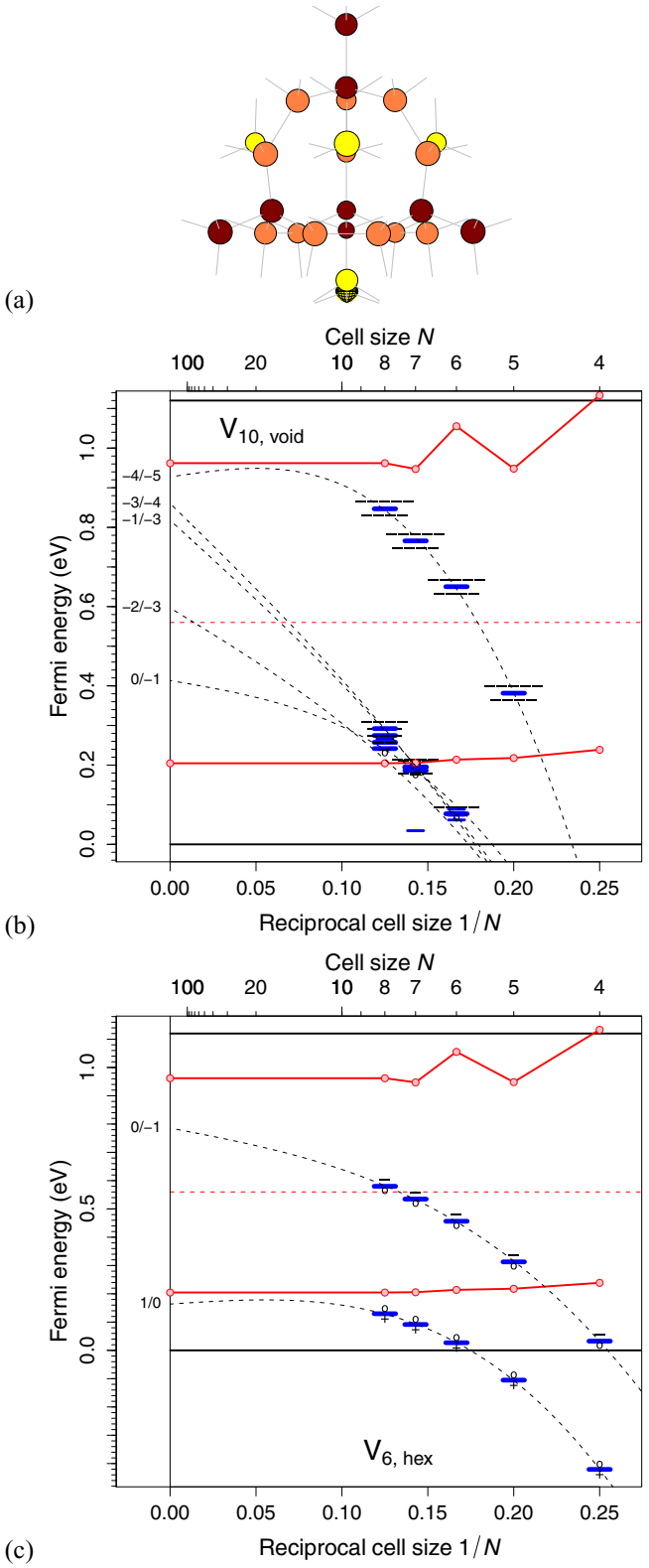

FIG. 10. (Color online) (a) $V_{10,\text{void}}$ relaxation after $V_{10,\text{void}}^0$ captures an electron. Threefold-coordinated atoms are yellow (light gray), dimerized atoms are orange (medium grey), and atoms with elongated bonds are brown (dark gray). The affected atom is at the bottom of the cluster, and its initial position is marked; other atoms hardly move. View along $(1\bar{2}1)$. (b) Cell-size convergence of $V_{10,\text{void}}$ gap levels; $dE_F = 0.19$ eV. (c) The same for $V_{6,\text{hex}}$. For the extrapolation method, see text.

TABLE VII. $E_{\rm form}$ of metastable V_n, $8 \leqslant n \leqslant 11$, at GGA VBM.

$7\times7\times7$	spin	$V_{8,\rm hex}$	$V_{9,\rm hex}$	$V_{10,\rm hex}$	$V_{11,\rm hex}$
0	no	13.131	14.460	15.473	16.537
−1	no	13.448	14.584	15.575	16.647
−2	no	13.833	14.774	15.743	16.824
−3	no	14.287	15.207	16.069	17.170

band; for this reason, we considered only uncharged and negatively charged clusters of this type. We find that the threefold-coordinated (3c) atoms capture excess electrons one by one. The excess electron in $V^-_{n,\rm void}$ goes to one dangling bond, making the affected atom relax inwards, while the other atoms are largely unaffected, apart from some elastic response [Fig. 10(a)]; this repeats (for another 3c atom) with each new electron added, up to the charge state $q = -4$. The decavacancy $V_{10,\rm void}$ behaves in the same way; this is in contrast to the breathing relaxation computed in Ref. [61] for $V_{10,\rm void}$ on electron capture. The cause for the difference may be that we did not impose any symmetry on the defect, either explicit or implicit; in fact, we always initialized the calculations with slightly nonsymmetric positions of atoms.

These differences notwithstanding, we confirm the observation in Ref. [61] that there exists strong lattice relaxation of the threefold-coordinated atoms in response to charge trapping and that this relaxation strongly reduces the electron-electron repulsion (effective Hubbard U or Anderson U). In 128-, 250-, and 432-site cells, this effect is strong enough to destabilize the charge states other than 0 and −4: the stable transition state is (0/−4). But in the $7\times7\times7$ cell (686 sites) the (−1/−3) level becomes stable, and in the $8\times8\times8$ cell (1024 sites) the Anderson U turns positive. Since the energy differences between the affected levels are small and the relaxation of the core depends on the cell size up to big cells, the extrapolation to infinity is not accurate enough to predict the ordering of the levels in the dilute limit [Fig. 10(a); see Sec. IV for more discussion on the extrapolation]. But it makes sense to expect that the cluster is neutral for $E_{\rm F}$ around midgap or that the energy of the neutral defect provides a good estimate of the $E_{\rm form}$ at midgap.

Eventually, we find that the void-type V_n clusters are either uncharged or negatively charged (we assume the same $dE_{\rm F}$ as for the hex clusters). We cannot exclude that a positively charged defect may exist in the dilute limit (and Fig. 10, taken together with the results of Ref. [61] for donor states, supports this view for $V_{10,\rm void}$), but we currently have no direct evidence for such a case.

In contrast to these complications, $V_{6,\rm hex}$ is well behaved [Fig. 10(b)] and so are the stable V_4 and V_5.

E. Binding and dissociation energies

The energies of uncharged clusters are compared in Fig. 11 to those in the literature [59,64–66]. All show the same trend: tendency to form clusters growing in size, enhanced stability of $V_{6,\rm hex}$ and $V_{10,\rm void}$, and relatively easy dissociation of V_7 and V_{11}. The literature data compare well to those obtained here for clusters of similar size; the older, less accurate density functional based tight binding (DFTB) and Harris functional

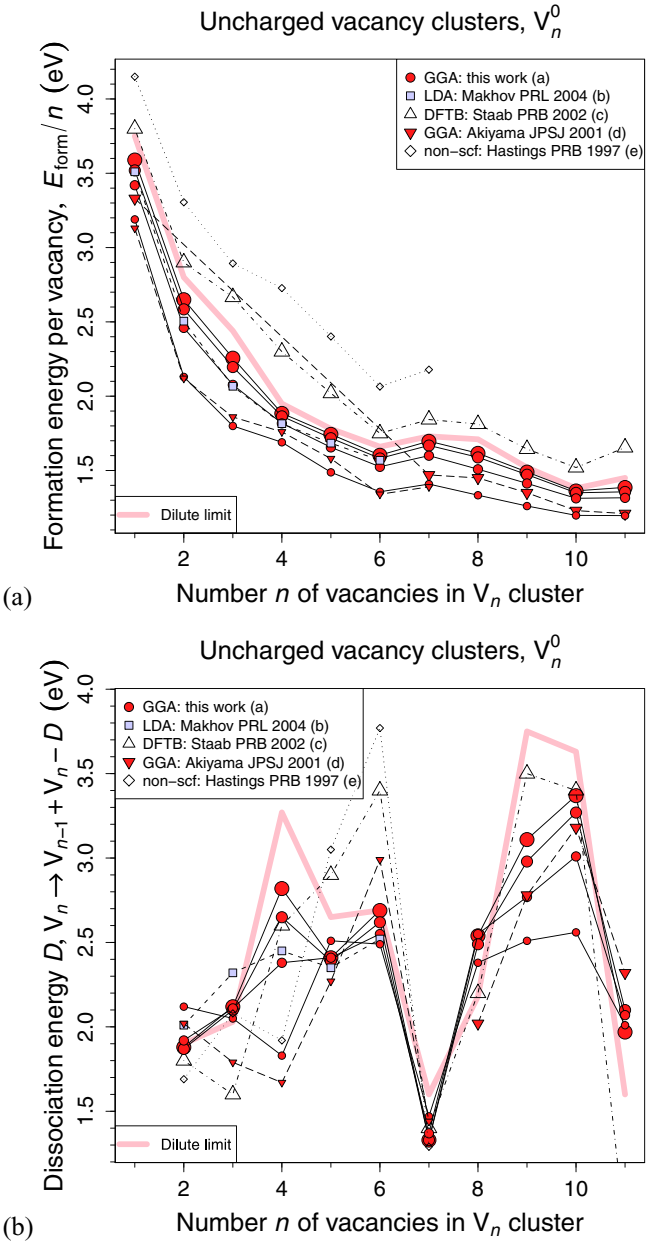

FIG. 11. (Color online) Cell-size convergence of (a) $E_{\rm form}/n$ and of (b) dissociation energy D in the reaction $V_n \rightarrow V_{n-1} + V_1 - D$ in cells with 128, 250, 432, and 868 sites. For the data, see Tables II (finite cells) and VIII (dilute limit). Compared to Refs. [59] (216 sites; labeled b), [64] (512 sites; c), [66] (64 and 216 sites; d), and [65] (64 sites; e). Larger cells are plotted with larger symbols.

calculations overestimate the absolute $E_{\rm form}$, but for most clusters the errors largely cancel out in D. The dissociation energy D of several clusters depends strongly on the cell size. Interestingly, this is most dramatic for the small, not for the large, $V_{n,\rm void}$ defects. The stability order for $n \leqslant 6$ is incorrect below 216 sites. The intriguing stability enhancement of V_2 occurs only in small cells. The dissociation energy of V_4 is the lowest in the hex group at fcc $4\times4\times4$ (128 sites), but this changes already with 250 sites, and from 432 sites on the dissociation of V_4 is even more difficult than the dissociation of V_6.

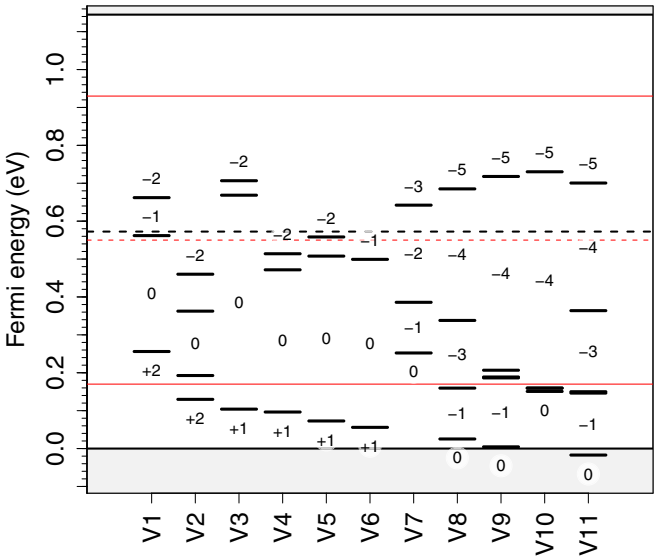

FIG. 12. (Color online) Stable levels in the $7 \times 7 \times 7$ cell (686 sites), adjusted by $dE_F = 0.19$ eV. GGA band edges are indicated as red thin solid lines, and the midgap energy is shown as a red thin dashed line. The estimated locations of true (many-particle) band edges and of the midgap are marked by black thick solid and dashed lines. The true VBM is dE_F below the GGA VBM.

F. Charge transition levels

Gap levels of stable V_n in the 686-site cell are collected in Fig. 12. The value of $dE_F = 0.19$ eV that was used to correct for the DFT band gap problem was obtained by the least-squares fit of the levels in the dilute limit (Sec. IV) to the experimental data for V_1 and V_3. As the GGA band gap is, for the pseudopotentials used in this work, equal to 0.76 eV, this yields nearly symmetric corrections to the VBM (0.19 eV) and to the CBM (0.18 eV). The midgap energy is thus nearly the same when estimated from the dE_F correction (0.19 eV) or when taking half of the GGA band gap of 0.76 eV (Fig. 12).

The lowest donor levels and, for $V_{n,\mathrm{void}}$, the lowest acceptor levels are under the GGA VBM even in the 686-site cell, as are the Kohn-Sham energies of the related orbitals. The resulting hybridization with the valence-band states and escape of the charge from the resonant state to the bulk may affect the extrapolation accuracy.

IV. EXTRAPOLATION TO DILUTE LIMIT

Figure 13 illustrates a common problem with extrapolation of formation energies to the dilute limit: in multiple cases, the correction magnitude noticeably depends on the core function applied, although the fit quality (measured by the sum of squared residuals σ) is for most functions comparable. For example, for V_1 we have

model:	$r2$	**r3**	$r4$	$r5$	$r6$	$r7$	$r8$	p	**e1**	$e2$
$\sigma/\sigma_{\mathrm{best}}$:	6.4	**1.7**	2.0	2.5	3.0	3.4	3.8	1.7	**1.0**	1.3

Poor performance of r2 prevails for all other defects and indicates that the Γ-related wave-function overlap effects [78] are less dramatic in fcc than in sc [78] cells.

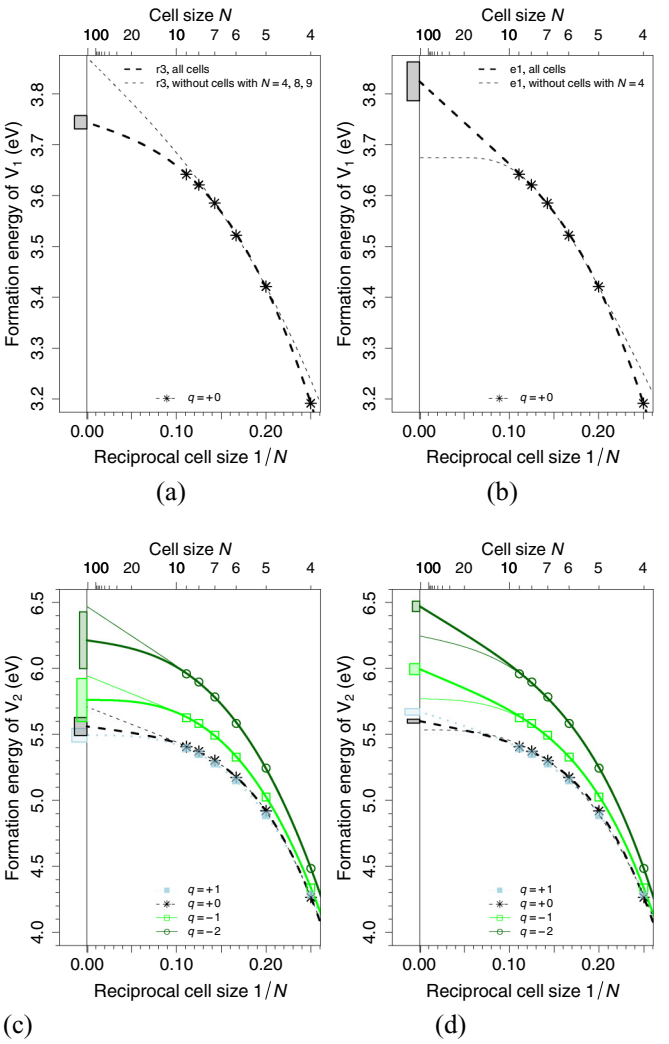

FIG. 13. (Color online) Extrapolated E_{form}. Boxes are the standard errors of the best fits. (a) The r3 and (b) the e1 models, applied to V_1^0 for all cells and with some cell sizes ignored. $E_{\mathrm{form}}(V_2^q)$ at $E_F = E_{\mathrm{VBM}}^{\mathrm{GGA}}$ for the two best (c) four-parameter and (d) three-parameter models.

The accuracy of $E_{\mathrm{form}}(\infty)$ may be assessed by comparing the estimates obtained for the best models when all data points are used and when the smallest and the largest cells are ignored. This yields $E_{\mathrm{form}}(V_1^0, \infty) = 3.75 \pm 0.08$ eV. Within the 3-parameter class, one obtains $E_{\mathrm{form}}(V_1^0, \infty) = 3.75 \pm 0.13$ eV. The average of the best fits (e1, e2, p, r3, r4) with all cell sizes gives $E_{\mathrm{form}}\, V_1^0, \infty = 3.80 \pm 0.05$ eV. Inaccuracies of this order are typical also for other defects, although details vary from case to case.

The short-range core models are only test functions, so errors may add up when gap levels are computed from the extrapolated $E_{\mathrm{form}}(\infty)$. But the levels are defined by differences in E_{form} [Eq. (3)]. The real cores of similar charge states are likely to be similar and to cancel out when two raw formation energies are subtracted from one another to compute the raw gap level (this may be the reason why the Γ point is so efficient for gap-level calculations). One can thus take the model that works best for $E_{\mathrm{form}}(V_n)$, remove the constraint

TABLE VIII. $E_{\rm form}(q)$ for V_n in the 686-sites cell and extrapolation to infinite cell size by the best "r*i*" model [Eq. (9)]; q_i is the charge state stable at $E_{\rm F} = \frac{1}{2}E_{\rm gap}^{\rm exp} - dE_{\rm F} = 0.38$ eV above GGA VBM (intrinsic Si). For $E_{\rm form}(V_n^0)/n$ plots, see Fig. 11.

n, r*i*	$E_{\rm form}(0)$ 686	$E_{\rm form}(0)$ ∞	q_i 686	$E_{\rm form}(q_i)$ 686	q_i ∞	$E_{\rm form}(q_i)$ ∞
1, r3	3.585	3.74 ± 0.1	0	3.585	0	3.74 ± 0.1
2, r4	5.301	5.60 ± 0.1	−2	5.104	0	5.60 ± 0.1
3, r4	6.994	7.17 ± 0.2	−2	6.919	0	7.17 ± 0.2
4, r5	7.541	7.81 ± 0.1	−2	7.452	0	7.81 ± 0.1
5, r4	8.722	8.91 ± 0.2	−2	8.670	0	8.91 ± 0.2
6, r5	9.624	9.95 ± 0.2	−1	9.563	0	9.95 ± 0.2
7, r4	11.880	12.12 ± 0.2	−2	11.572	−1	12.09 ± 0.2
8, r4	12.928	13.69 ± 0.5	−4	12.393	−1	13.35 ± 0.3
9, r4	13.410	13.67 ± 0.1	−4	12.855	−1	13.26 ± 0.2
10, r4	13.625	13.84 ± 0.1	−4	13.216	−1	13.64 ± 0.3
11, r6	15.248	15.66 ± 0.5	−4	14.671	0	15.66 ± 0.5

on A, and then fit it to V_n *levels*. We find that this works much better than extrapolating the levels by subtracting two $E_{\rm form}(\infty)$.

Using this method, we obtained the levels of stable V_n in the dilute limit (Fig. 14). Here, we took the average of boundary cases (best model and the second-best model or ignored cells of extreme size) shown as dotted lines, and we set $dE_{\rm F}$ to 0.18 eV (the average of best $dE_{\rm F}$ for V_1 and V_3, with V_2 ignored because of its exotic bonding).

One can take advantage of the error cancellation in levels and improve the estimate of some $E_{\rm form}V_n^{q1}$ by applying Eq. (3) to a more accurate $E_{\rm form}V_n^{q2}$ and a more accurate level (q_1/q_2). Table VIII collects the cluster formation energies $E_{\rm form}V_n^q$, estimated for the intrinsic Si. The V_n^0 energies, expressed per vacancy, $E_{\rm form}/n$, and as dissociation energies, $D_n = nE_{\rm form}(V_1^0) - E_{\rm form}(V_n^0)$, are indicated in Fig. 11 by thick pale lines.

Given Table VIII, the dissociation energy D of the V_2^0 cluster into two V_1^0 becomes $DV_2^0 = 1.9 \pm 0.3$ eV. According to experiment [40], $D(V_2^0)$ should be least 1.6 eV, which is consistent with this estimate.

V. SOME IMPLICATIONS

A. Simulation of vacancy clustering

During thermal quenching, vacancy clustering is expected to happen around 600–700 °C (Figs. 4 and 6 in Ref. [67]). By dopant concentrations in Si wafers of up to about 10^{18} cm^{-3}, the Fermi level is then in the midgap region [84]. To assess relative V_n concentrations at such conditions, one should therefore know the V_n^q formation energies for q that are stable around midgap.

In the 686-site cell, only V_1 can be uncharged around midgap (Fig. 12). In the diluted limit (Fig. 14), most acceptor states move upwards, so that more clusters may be in the neutral state (V_1–V_6, V_{10}, V_{11}). This means that, apart from underestimating the formation energy of each charge state separately, calculations in finite cells underestimate the formation energies of overcharged clusters (like of V_2 or of V_8) by an additional amount, equal to the excess charge times the energy difference between the midgap energy and the acceptor level [Eq. (3)]. This improves when only uncharged defects are considered in finite cells, as is usually done. The influence of charging tends to cancel out in dissociation energies (Fig. 15).

In our previous work, we adapted for kinetic calculations of V clustering the energies from $5 \times 5 \times 8$ cells [67]. This resulted in decreased stability of V_n for $n < 6$, and the V_6 clusters turned out to be the dominant species after RTA at 1250 °C and during anneals at temperatures lower than 800 °C [67]. With the energies extrapolated to the diluted limit in this paper, V_6 remains dominant after RTA because the process is fast and the low stability of the V_7 cluster results in a kinetic barrier. But a preceding anneal below 1000 °C would make the V_{10} cluster the dominant species. Increased V_9 stability (Fig. 15) does not hinder the growth of V_{10} because the kinetic barrier is not too high and because V_{11} still dissociates easily.

B. Vacancy clusters in irradiated samples

When vacancies are produced by particle irradiation, their clustering takes place at temperatures much lower than during thermal quenching. Metastable structures are therefore likely to be seen in the measured spectra.

In Sec. III D, we mentioned three metastable clusters that have been identified by EPR in irradiated Si: the ⟨110⟩-planar chain [55,59,64] ($V_{4,110}^0$) assigned to the P3 spectrum [55], the ⟨111⟩-planar chain [55] ($V_{4,\rm void}^-$) assigned to the A3 spectrum [55], and the nonplanar $V_{5,\rm void}^-$ assigned to the P-1 spectrum [68]. We confirm that these defects are EPR active in these charge states. But during thermal anneals, the P3 spectrum is replaced by the A3 spectrum at $T_{\rm o} = 170$ °C. Hence, $E_{\rm F}$ for which $V_{4,110}^0$ is *less stable* than $V_{4,\rm void}^-$ exists. Conversely, Table VI states that $V_{4,110}$ is *more stable* than $V_{4,\rm void}$ for all charge states (i.e., also for all $E_{\rm F}$). If the assignment from EPR is correct, this discrepancy may mean that the contribution of vibrational entropy to the free energy of $V_{4,\rm void}$ at $T_{\rm o} = 170$ °C is large enough to stabilize this cluster with respect to $V_{4,110}$. Such a stability reversal would require the entropy term $T_{\rm o}$ to be at least 0.2–0.3 eV, or less than the vibrational $ST_{\rm o}$ of a single vacancy [25]; hence, this hypothesis is not implausible.

The EPR signal from $V_{4,\rm void}^-$ disappears at 350 °C. No V_4 signal is seen above this temperature [55]. This may be due to the conversion of $V_{4,\rm void}^-$ into $V_{4,\rm FFC}^0$. Namely, the only charge state of $V_{4,\rm void}$ that is active in EPR is $V_{4,\rm FFC}^-$, and it can exist only at Fermi levels about 0.15 eV higher than the upper limit for $V_{4,\rm void}^-$ (Fig. 16).

C. Implications for defect simulation

One of the important problems in theoretical defect physics is the underestimate of the band gap by the DFT with exchange and correlation energy described be LDA or GGA formulas [93,116]. We addressed it applying to the GGA band edges an *a posteriori* correction estimated by comparing the calculated gap levels of V_1 and V_3 to experiment (the marker method). But the results cease then to be fully "from first principles"; worse, it is unclear which marker to choose because marker performance is affected by the character of the defect wave function.

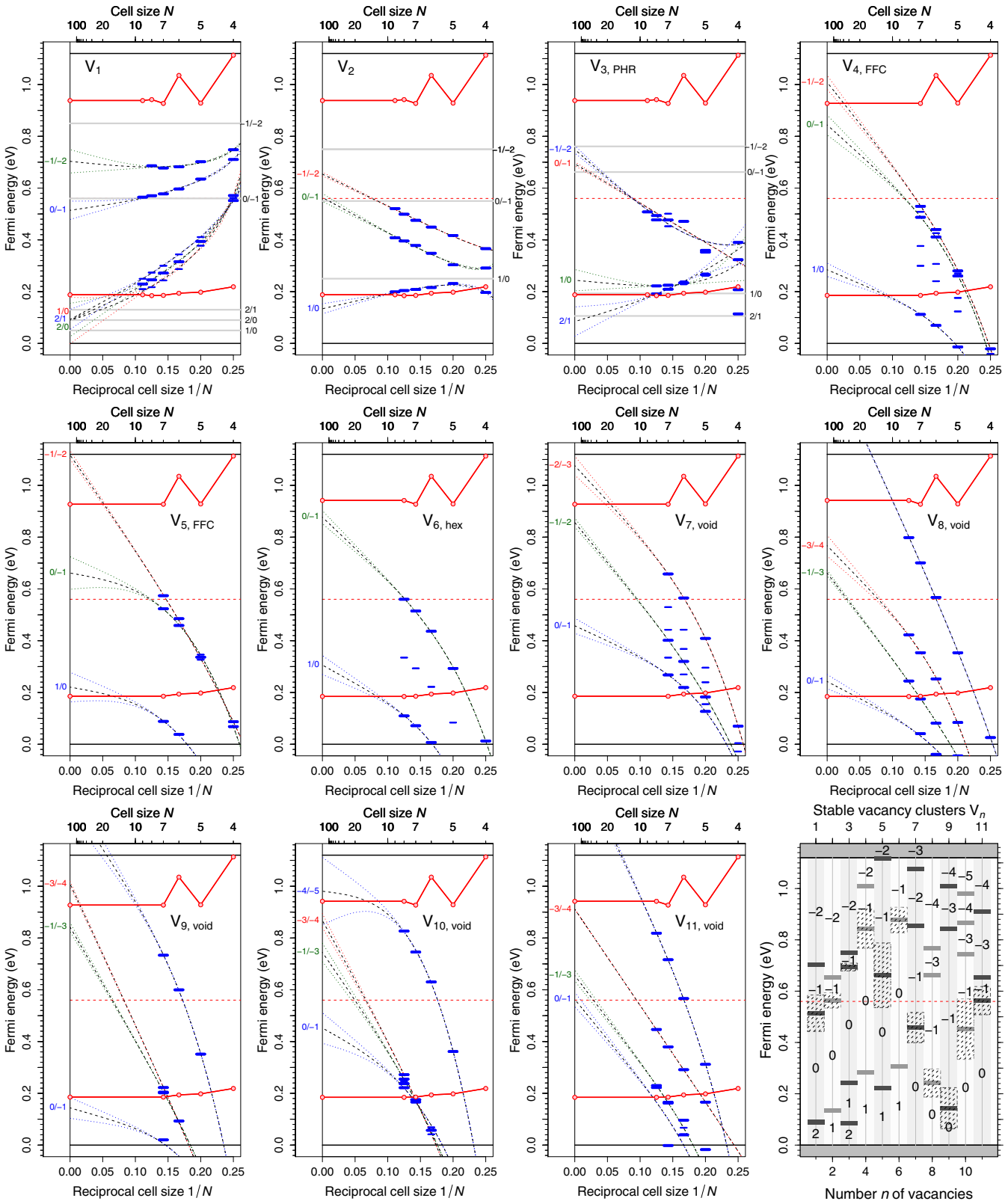

FIG. 14. (Color online) Stable (long dashes) and some unstable (short dashes) levels of V_n, adjusted by $dE_F = 0.19$ eV. The D_{3d} state of V_1 is ignored (see text). Extrapolation is done with no constraint on A. Shaded boxes show the estimated uncertainty of $(0/-1)$ energies. The interconnected points are the GGA band extrema. Gray lines mark the DLTS data [5,40,46,56,57,108].

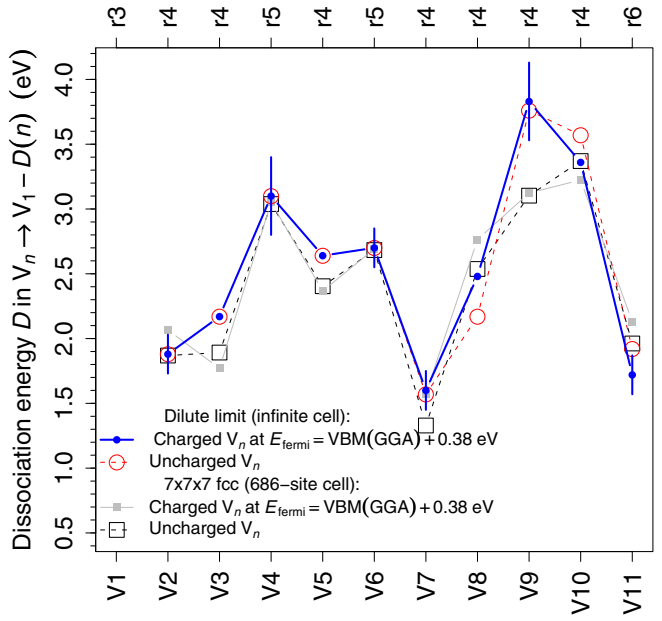

FIG. 15. (Color online) V dissociation energies computed with the $7 \times 7 \times 7$ cell (686 sites) and extrapolated to infinite supercells by the best ri model [Eq. (9)] for each defect. See Table VIII.

There exist truly first-principles methods to overcome the band gap problem that are applicable in practice also for calculations with the 64-site cell or even with over hundred atoms. Notably, one can employ the zero-order GW approach (G_0W_0) to compute quasiparticle energies [117–120], or one can mix in some Hartree-Fock exchange by using hybrid potentials [121–125]. Both tools are implemented in the leading *ab initio* packages, including QUANTUM ESPRESSO, and are becoming increasingly popular. Nevertheless, their computational costs are markedly higher than those of the usual LDA or GGA DFT, so that saving the numerical effort on the supercell size and on the k-point sampling becomes a priority.

We find that the performance of the (computationally inexpensive) [98] Γ sampling is better than its reputation. The wave-function overlap decays with the cell size L not slower than L^{-3} or L^{-4}, at least for fcc. With denser k grids, clusters do not relax much better. To the contrary, the bulk of the additional relaxation produced by sampling the BZ interior away from Γ is due to nonphysical occupation of states in partially filled

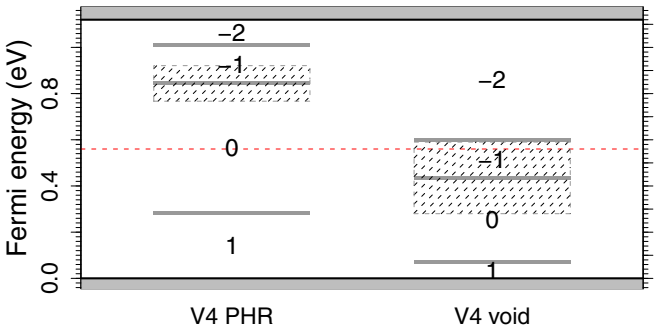

FIG. 16. (Color online) Levels of $V_{4,\mathrm{hex}}$ (stable) and $V_{4,\mathrm{void}}$ (A3). Dilute limit. Boxes indicate the inaccuracy of level determination.

defect bands [97]. When the cells are big enough to make the occupation effect insignificant (a 250-site cell may suffice), the relaxation effect becomes of little significance as well (3 meV for V_1, 17 meV for V_2). The major profit from no-Γ grids is a reduction in the cell-size dependence of the energies. The interaction of defect wave functions in the superlattice is of bonding type at Γ (the phase is the same in each supercell), while other k points add some antibonding character, partially averaging the interaction away.

A good strategy is thus to perform the relaxation at Γ and then recalculate the energies with another k-point set, but at fixed atomic positions. Still, we observed that doing this with the smallest cells (like the 128-site fcc cell) leads to more confusion than help: the energy–cell-size dependence becomes nonmonotonous, and the extrapolation may come out completely wrong if, e.g., only two or three cell sizes are examined.

Extrapolation to the dilute limit would be greatly facilitated when at least a part of the interaction between the defects in the superlattice is rigorously removed. There exists a recipe for doing this for the electrostatic term [80,81]. Yet one must verify if the cell is big enough for the nonelectrostatic terms to disappear. A reliable test requires at least three cell sizes. If the cells are big enough, the r3 model removes the electrostatics as well.

VI. SUMMARY AND CONCLUSIONS

We reported results of *ab initio* GGA calculations for formation energies E_{form} of vacancy complexes V_n^q in Si ($1 \leqslant m \leqslant 11$). Supercell size convergence was examined. We discussed implications for the simulation of vacancy clustering after thermal quenching (Sec. V A), for the interpretation of DLTS and EPR spectra in irradiated Si (Sec. V B), and for the strategy for efficient simulation of defects with methods beyond the standard (LDA or GGA) DFT (Sec. V C). In particular, we found the following:

(1) The acceptor levels of V_1 extrapolated from SiGe are well predicted if one assumes that the presence of Ge destabilizes the D_{3d} state (Sec. III A).

(2) The stability order of small ($n \leqslant 6$) clusters is incorrect in supercells below 216 sites (Sec. III E).

(3) V_3 to V_6 form rings built on the basis of the sixfold Si rings. In intrinsic Si they are uncharged. Clusters larger than V_7 and based on interwoven sixfold rings are metastable (Secs. III C and III D).

(4) Vibrational entropy may be needed to predict the stability sequence of metastable V_4 seen in EPR, but the behavior of the A3 and P3 spectra of V_4 is well described by E_{form} alone (Sec. V B).

(5) V_n dissociation energies strongly depend on the cell size, but the prediction that V_6 dominates after thermal quenching is unaffected (Sec. V A).

(6) V_6 may have two levels in the gap (Sec. III D).

(7) For $n \geqslant 7$, octahedral voids are stable, with $q = -1$ in intrinsic Si. The excess electron is trapped by one of the atoms in the void corners (Sec. III D).

(8) $V_{n,\mathrm{void}}^{2-}$ states may be unstable (Fig. 14).

(9) For the GGA pseudopotentials used here (Table I), the marker method yields a VBM correction of 0.19 eV, which

places the corrected and uncorrected midgaps at nearly the same energy (Sec. III F).

(10) With the fcc cells, Γ sampling is best suited for atomic relaxation and level calculation. Additional calculation at fixed atomic positions done outside Γ reduces the cell-size dependence of E_{form}, but with small cells caution is advised (Sec. V C).

(11) At Γ, the extrapolation of E_{form} to the dilute limit ($r = \infty$) works well with test functions of the type $E_\infty - |A|r^{-1} + Br^{-m}$, $3 \leqslant m \leqslant 6$. The r^{-1} term must be negative for any BZ grid used (Sec. II).

(12) Gap levels are best extrapolated directly, not from the extrapolated $E_{\text{form}}(\infty)$ (Sec. IV).

(13) What marker level should be used depends on the relative admixture of VB and CB states to the levels to be corrected (Secs. II, III A–III C, and IV).

ACKNOWLEDGMENTS

Ab initio calculations were performed at the Jülich Supercomputing Centre (JSC), Jülich, Germany. We gratefully acknowledge the computing time granted by the John von Neumann Institute for Computing (NIC) for NIC project fho06. Ab initio data were analyzed and the results were plotted by scripts written in the free-domain R *language environment* [106].

APPENDIX A: ORTHORHOMBIC VS FCC CELLS

Supercells with varying Wigner-Seitz shape may be, in principle, employed in the same study to increase the amount of information for the extrapolation scheme [79]. We tested sc, fcc, bcc, and orthorhombic cells. We quote the results for fcc cells, unless specified otherwise.

The fcc lattice vectors are $\frac{ma_0}{2}(1,0,1)$, $\frac{ma_0}{2}(1,1,0)$, and $\frac{ma_0}{2}(0,1,1)$, where m is the periodicity and a_0 is the cubic lattice constant. These cells have the full symmetry of the bulk. The primitive fcc cell has two atoms, the distance between defects in an $m \times m \times m$ cell is $d = \frac{\sqrt{2}}{2}ma_0$.

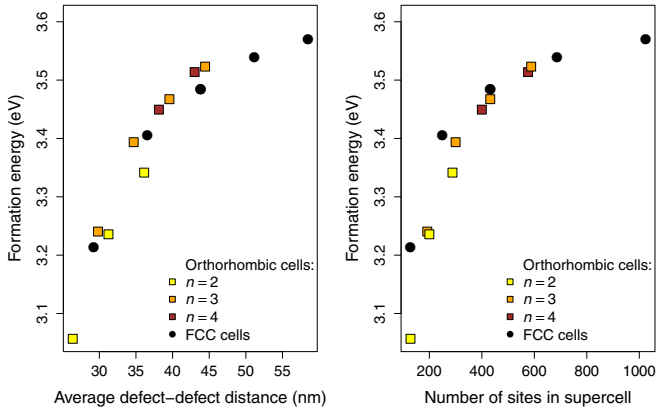

FIG. 17. (Color online) Formation energy of uncharged vacancy computed for fcc and orthorhombic Bravis lattices. The SI.PBE-N-VAN pseudopotential from QUANTUM ESPRESSO [85] distribution was used. Dependence (a) on the average distance between defects, defined as in the text, and (b) on the supercell volume.

We also tried orthorhombic cells built as $m \times m \times n$ multiples of the four-atom cell with the translation vectors of $a_0\sqrt{2}(1,1,0)$, $a_0\sqrt{2}(\bar{1},1,0)$, and $a_0(0,0,1)$. The distance between the neighboring defects is $d_{xy} = \frac{\sqrt{2}}{2}ma_0$ along (110) and $(\bar{1},1,0)$ and $d_z = na_0$ along (001). There are four d_{xy} and two d_z nearest neighbors; hence, the average distance between defects is, in this case, $d = \frac{2}{3}d_{xy} + \frac{1}{3}d_y$.

As illustrated by Fig. 17 for the example of V_1^0, E_{form} converges faster when fcc rather than orthorhombic cells are used. More importantly, the shapes of orthorhombic cells are many, and this is reflected in more complex size dependence than for fcc cells. The fcc family is thus better suited for a systematic convergence study.

APPENDIX B: CONVERGENCE OF STRUCTURAL RELAXATION

By inspection of the dependence of A (which measures the asymptotic part of the correction) on the charge state q one may estimate the contribution of elastic forces to the asymptotic behavior of E_{form} [Eq. (12)]. All extrapolation

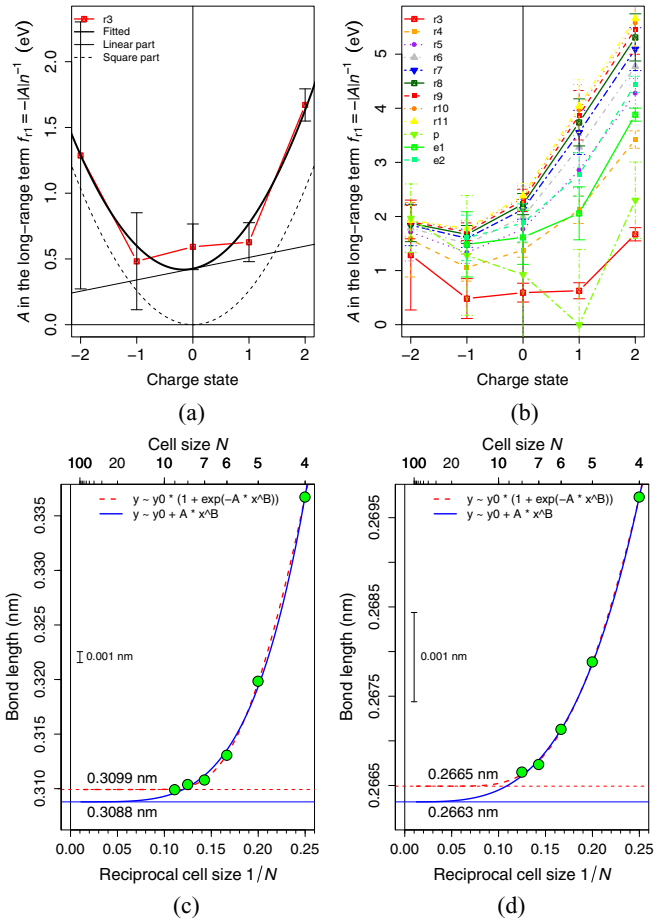

FIG. 18. (Color online) Relaxation of V_1. Dependence of the long-range interaction strength $-|A|r^{-1}$ [Eq. (8)] on the charge state q: (a) $-|A|(q)$ decomposed into linear (elastic) and square (elastic, electrostatic) terms for the r3 model and (b) $-|A|(q)$ for well-performing models. (c) and (d) Structure. (c) Dimer bond length of V_1^0; point group (pg) symmetry is D_{2d} in all cells. (d) Bond length in V_1^{2-}; pg symmetry is D_{3d} in all cells.

models yield a term linear in q [Figs. 18(a) and 18(b)]. This indicates that this contribution is detectable, at least in the fcc superlattice. This is in line with the observation that the strain field of V_1 expands along ⟨110⟩; that is, it coincides with the supercell translation vectors [24]. The elastic contribution was reported to be weak for V_1 in bcc [24] and sc cells [28]. For fcc, the pure q^2 dependence remains within the error bars of the r3 model, but this is not so for more precise models (p, e1). Since the V_1 atomic core practically converges in finite cells [Figs. 18(c) and 18(d)], the asymptotic elastic contribution stems from the bulk, as expected from the theory of elasticity.

[1] J. W. Corbett and G. D. Watkins, Phys. Rev **138**, A555 (1965).
[2] G. D. Watkins, in *Lattice Defects in Semiconductors 1974*, edited by F. A. Huntley, Institute of Physics Conference Series Vol. 23 (Institute of Physics, London, 1975), p. 1.
[3] J. W. Corbett and J. C. Bourgoin, in *Semiconductors and Molecular Crystals*, edited by J. H. Crawford, Jr., and L. M. Slifkin (Plenum, New York, 1975), Vol. 2, p. 1.
[4] G. D. Watkins and J. R. Troxell, Phys. Rev. Lett. **44**, 593 (1980).
[5] G. D. Watkins, in *Deep Centers in Semiconductors*, edited by S. T. Pantelides (Gordon and Breach, New York, 1986), p. 147.
[6] G. D. Watkins, Mater. Res. Soc. Symp. Proc. **469**, 139 (1997).
[7] M. Jacob, P. Pichler, H. Ryssel, and R. Falster, J. Appl. Phys. **82**, 182 (1997).
[8] G. D. Watkins, Mater. Sci. Semiconduct. Process. **3**, 227 (2000).
[9] G. A. Baraff and M. Schlüter, Phys. Rev. Lett. **41**, 892 (1978).
[10] J. Bernholc, N. O. Lipari, and S. T. Pantelides, Phys. Rev. Lett. **41**, 895 (1978).
[11] G. A. Baraff and M. Schlüter, Phys. Rev. B. **19**, 4965 (1979).
[12] J. Bernholc, N. O. Lipari, and S. T. Pantelides, Phys. Rev. B. **21**, 3545 (1980).
[13] M. J. Puska, O. Jepsen, O. Gunnarsson, and R. M. Nieminen, Phys. Rev. B **34**, 2695 (1986).
[14] G. A. Baraff, E. O. Kane, and M. Schlüter, Phys. Rev. Lett. **43**, 956 (1979).
[15] G. A. Baraff, E. O. Kane, and M. Schlüter, Phys. Rev. B **21**, 5662 (1980).
[16] R. Car, P. J. Kelly, A. Oshiyama, and S. T. Pantelides, Phys. Rev. Lett. **52**, 1814 (1984).
[17] M. J. Puska, J. Phys. Condens. Matter **1**, 7347 (1989).
[18] P. J. Kelly and R. Car, Phys. Rev. B **45**, 6543 (1992).
[19] M. A. Roberson and S. K. Estreicher, Phys. Rev. B **49**, 17040 (1994).
[20] O. Sugino and A. Oshiyama, Phys. Rev. Lett. **68**, 1858 (1992).
[21] M. J. Puska, S. Pöykkö, M. Pesola, and R. M. Nieminen, Phys. Rev. B **58**, 1318 (1998).
[22] R. M. Nieminen and M. J. Puska, in *Properties of Crystalline Silicon*, edited by R. Hull (INSPEC, London, 1999), p. 309.
[23] J. Dabrowski, Solid State Phenom. **71**, 23 (2000).
[24] M. I. J. Probert and M. C. Payne, Phys. Rev. B **67**, 075204 (2003).
[25] O. K. Al-Mushadani and R. J. Needs, Phys. Rev. B. **68**, 235205 (2003).
[26] P. A. Schultz, Phys. Rev. Lett. **96**, 246401 (2006).
[27] A. F. Wright and N. A. Modine, Phys. Rev. B **74**, 235209 (2006).
[28] A. F. Wright, Phys. Rev. B **74**, 165116 (2006).
[29] M. G. Ganchenkovaa, L. E. Oikkonen, V. A. Borodin, S. Nicolaysen, and R. M. Nieminen, Mater. Sci. Eng. B **159-160**, 107 (2009).
[30] F. Corsetti and A. A. Mostofi, Phys. Rev. B **84**, 035209 (2011).
[31] P. Fahey, G. Barbuscia, M. Moslehi, and R. W. Dutton, Appl. Phys. Lett. **46**, 784 (1985).
[32] P. M. Fahey, P. B. Griffin, and J. D. Plummer, Rev. Mod. Phys. **61**, 289 (1989).
[33] S. T. Dunham and C. D. Wu, J. Appl. Phys. **78**, 2362 (1995).
[34] O. Pankratov, H. Huang, T. Diaz de la Rubia, and Ch. Mailhiot, Phys. Rev. B **56**, 13172 (1997).
[35] *Identification of Defects in Semiconductors*, edited by M. Stavola (Academic, San Diego, 1998).
[36] M. Sprenger, S. H. Muller, E. G. Sieverts, and C. A. J. Ammerlaan, Phys. Rev. B **35**, 1566 (1987).
[37] P. W. Anderson, Phys. Rev. Lett. **34**, 953 (1975).
[38] J. W. Corbett, G. D. Watkins, and R. S. McDonald, Phys. Rev. **135**, A1381 (1964).
[39] G. D. Watkins and J. W. Corbett, Phys. Rev. **134**, A1359 (1964).
[40] G. D. Watkins and J. W. Corbett, Phys. Rev **138**, A543 (1965).
[41] L. J. Cheng, J. C. Corelli, J. W. Corbett, and G. D. Watkins, Phys. Rev **152**, 761 (1966).
[42] C. A. J. Ammerlaan and G. D. Watkins, Phys. Rev. B **5**, 3988 (1972).
[43] A. O. Evwaraye and E. Sun, J. Appl. Phys. **47**, 3776 (1976).
[44] E. G. Sieverts, S. H. Muller, and C. A. J. Ammerlaan, Phys. Rev. B **18**, 6834 (1978).
[45] B. G. Svensson, B. Mohadjeri, A. Hallén, J. H. Svensson, and J. W. Corbett, Phys. Rev. B **43**, 2292 (1991).
[46] H. Kauppinen, C. Corbel, J. Nissilä, K. Saarinen, and P. Hautojärvi, Phys. Rev. B **57**, 12911 (1998).
[47] S. Öğüt and J. R. Chelikowsky, Phys. Rev. Lett. **83**, 3852 (1999).
[48] S. Öğüt and J. R. Chelikowsky, Phys. Rev. B **64**, 245206 (2001).
[49] M. Saito and A. Oshiyama, Phys. Rev. Lett. **73**, 866 (1994).
[50] H. Seong and L. J. Lewis, Phys. Rev. B **53**, 9791 (1996).
[51] M. Pesola, J. von Boehm, S. Pöykkö, and R. M. Nieminen, Phys. Rev. B **58**, 1106 (1998).
[52] A. L. S. Ferreira-Resende, Ph.D. thesis, University of Exeter, 1999.
[53] R. R. Wixom and A. F. Wright, Phys. Rev. B **74**, 205208 (2006); **75**, 199901(E) (2007).
[54] J. I. Iwata, K. Shiraishi, and A. Oshiyama, Phys. Rev. B **77**, 115208 (2008).
[55] Y.-H. Lee and J. W. Corbett, Phys. Rev. B **9**, 4351 (1974).
[56] V. P. Markevich, A. R. Peaker, S. B. Lastovskii, L. I. Murin, J. Coutinho, A. V. Markevich, V. J. B. Torres, P. R. Briddon, L. Dobaczewski, E. V. Monakhov, and B. G. Svensson, Phys. B (Amsterdam, Neth.) **404**, 4565 (2009).
[57] V. P. Markevich, A. R. Peaker, B. Hamilton, S. B. Lastovskii, L. I. Murin, J. Coutinho, V. J. B. Torres, L. Dobaczewski, and G. B. Svensson, Phys. Status Solidi A **208**, 568 (2011).

[58] J. Coutinho, V. P. Markevich, A. R. Peaker, B. Hamilton, S. B. Lastovskii, L. I. Murin, B. J. Svensson, M. J. Rayson, and P. R. Briddon, Phys. Rev. B **86**, 174101 (2012).

[59] D. V. Makhov and L. J. Lewis, Phys. Rev. Lett. **92**, 255504 (2004).

[60] S. K. Estreicher, J. L. Hastings, and P. A. Fedders, Appl. Phys. Lett. **70**, 432 (1997).

[61] K. Uchida and A. Oshiyama, J. Phys. Soc. Jpn. **79**, 093711 (2010).

[62] D. J. Chadi and K J. Chang, Phys. Rev. B **38**, 1523 (1988).

[63] T. Akiyama, A. Oshiyama, and O. Sugino, J. Phys. Soc. Jpn. **67**, 4110 (1998).

[64] T. E. M. Staab, A. Sieck, M. Haugk, M. J. Puska, T. Frauenheim, and H. S. Leipner, Phys. Rev. B **65**, 115210 (2002).

[65] J. L. Hastings, S. K. Estreicher, and P. A. Fedders, Phys. Rev. B **56**, 10215 (1997).

[66] T. Akiyama and A. Oshiyama, J. Phys. Soc. Jpn. **70**, 1627 (2001).

[67] G. Kissinger, J. Dabrowski, and D. Kot, Jpn. J. Appl. Phys. **53**, 05FJ06 (2014).

[68] Y.-H. Lee and J. W. Corbett, Phys. Rev. B **8**, 2810 (1973).

[69] S. W. de Leeuw, J. W. Perram, and E. R. Smith, Proc. R Soc. London, Ser. A **373**, 27 (1980).

[70] M. Leslie and M. J. Gillan, J. Phys. C **18**, 973 (1985).

[71] G. Makov and M. C. Payne, Phys. Rev. B **51**, 4014 (1995).

[72] J. Lento, J.-L. Mozos, and R. M. Nieminen, J. Phys.: Condens. Matter **14**, 2637 (2002).

[73] A. Höglund, C. W. M. Castleton, and S. Mirbt, Phys. Rev. B **72**, 195213 (2005).

[74] C. W. M. Castleton, A. Höglund, and S. Mirbt, Phys. Rev. B **73**, 035215 (2006).

[75] S. Lany and A. Zunger, Phys. Rev. B **78**, 235104 (2008).

[76] R. M. Nieminen, Modell. Simul. Mater. Sci. Eng. **17**, 084001 (2009).

[77] S. Lany and A. Zunger, Modell. Simul. Mater. Sci. Eng. **17**, 084002 (2009).

[78] C. W. M. Castleton, A. Höglund, and S. Mirbt, Modell. Simul. Mater. Sci. Eng. **17**, 084003 (2009).

[79] N. D. M. Hine, K. Frensch, W. M. C. Foulkes, and M. W. Finnis, Phys. Rev. B **79**, 024112 (2009).

[80] Ch. Freysoldt, J. Neugebauer, and Ch. G. Van de Walle, Phys. Rev. Lett **102**, 016402 (2009).

[81] Ch. Freysoldt, J. Neugebauer, and Ch. G. Van de Walle, Phys. Status Solidi B **248**, 1067 (2011).

[82] S. T. Pantelides, Rev. Mod. Phys. **50**, 797 (1978).

[83] A. Bongiorno, L. Colombo, and T. Diaz De la Rubia, Europhys. Lett. **43**, 695 (1998).

[84] S. M. Sze, *Physics of Semiconductor Devices* (Wiley, New York, 1981).

[85] P. Giannozzi *et al.*, J. Phys. Condens. Matter **21**, 395502 (2009).

[86] J. P. Perdew, K. Burke, and M. Ernzerhof, Phys. Rev. Lett. **77**, 3865 (1996).

[87] The custom pseudopotentials SI.PBE-5170 and SI.PBE-5133 are PBE pseudopotentials generated with the LDX1 utility of the QE distribution, using the Rappe-Rabe-Kaxiras-Joannopoulos (RRKJ) recipe [126]. The all-electron potential was pseudized. The valence configuration was $3s^{1.7}3p^{2.3}$, and the s and p cutoff radii were set to 2.32 a.u., while the local potential cutoff radius was set to $rcloc = 2.15$ a.u for the SI.PBE-5170 pseudopotential and to $rcloc = 2.20$ a.u. for the SI.PBE-5133 pseudopotential.

[88] Ph. Haas, F. Tran, and P. Blaha, Phys. Rev. B **79**, 085104 (2009).

[89] T. Motooka and T. Uda, in Handbook of Silicon Based MEMS: Materials and Technologies, part 2, *Modeling in MEMS*, edited by M. Tilli, T. Motooka, V.-M. Airaksinen, S. Franssila, M. Paulasto-Kröckel, and V. Lindroos (Elsevier, London, 2015), p. 241; M. Ganchenkova and R. M. Nieminen, *ibid.*, p. 253.

[90] W. C. O'Mara, R. B. Herring, and L. P. Hunt, Handbook of Semiconductor Silicon Technology (Noyes, Park Ridge, NJ, 1990).

[91] M. A. Hopcroft, W. D. Nix, and T. W. Kenny, J. Microelectromech. Syst. **19**, 229 (2010).

[92] J. F. Janak, Phys. Rev. B **18**, 7165 (1978).

[93] R. W. Godby, M. Schlüter, and L. J. Sham, Phys. Rev. Lett. **56**, 2415 (1986).

[94] A. Resende, R. Jones, S. Öberg, and P. R. Briddon, Phys. Rev. Lett. **82**, 2111 (1999).

[95] J.-W. Jeong and A. Oshiyama, Phys. Rev. B **64**, 235204 (2001).

[96] J. P. Goss, P. R. Briddon, S. J. Sque, and R. Jones, Diamond Relat. Mater. **13**, 684 (2004).

[97] A partial remedy would be to enforce equal distribution of defect charge in the k space [26, 127, 104]. But then we do not compute the ground state of the electron subsystem defined by the k-point set. In addition, if the defect has degenerate or nearly degenerate gap states at Γ that are still noticeably dispersive, we may bias the average energy towards an arbitrary binding combination of the wave functions of defects in the supercell, and we may introduce bias to the atomic relaxation mode.

[98] The wave functions at Γ are real, but elsewhere in the interior of the Brillouin zone they are complex.

[99] At Γ, the lattice constant is converged to 0.13% error for the smallest $4 \times 4 \times 4$ cell (128 sites), to 0.04% error for the $5 \times 5 \times 5$ cell (250 sites), and to 0.00% error for the $6 \times 6 \times 6$ (432 sites) cell. The convergence errors for the bulk modulus are 2.2% (128 sites), 0.0% (250 sites), and 0.0% (432 sites), and for the pressure derivative of the bulk modulus they are -4%, 0%, and -1%, respectively.

[100] L. D. Landau and E. M. Lifshitz, *Theory of Elasticity* (Pergamon, Oxford, 1986).

[101] B. Benedikt, M. Lewis, and P. Rabgsawamy, Comput. Mater. Sci. **37**, 380 (2006).

[102] Since the energy of the strain, electrostatic, or gravitational field is defined as the work done from the "no interaction" (i.e., "no field") to the "full interaction" case, the the sign of the energy density in the case of fields surrounding a single object or caused by interaction between identical objects must always be positive in electrostatic and elastic interactions (repulsive interactions). In the case of gravitational interactions the sign is, of course, negative.

[103] This is done to avoid spurious interactions between the background charge and the electrons.

[104] C. G. Van de Walle and J. Neugebauer, J. Appl. Phys. **95**, 3851 (2004).

[105] The PORT library, http://www.netlib.org/port/.

[106] R Core Team, *R: A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, Vienna, 2013); http://www.R-project.org/.

[107] H. J. Monkhorst and J. D. Pack, Phys. Rev. B **13**, 5188 (1976).

[108] A. Nylandsted Larsen, A. Bro Hansen, and A. Mesli, Mater. Sci. Eng. B **154-155**, 85 (2008).
[109] C. V. Budtz-Jørgensen, P. Kringhøj, A. Nylandsted Larsen, and N. V. Abrosimov, Phys. Rev. B **58**, 1110 (1998).
[110] A. Nylandsted Larsen (private communication).
[111] Interestingly, no negative-U behavior has been predicted by DFT for acceptors states of a single vacancy in Ge [128].
[112] This inaccuracy would decrease to about 3 meV/vacancy when the energy cutoff is increased from 10 to 12 Ry. For 15 Ry it would drop below 1 meV, and for 8 Ry it would exceed 50 meV. Given the high number of structures and charge states treated here and the intention to reach cells that are as big as practicable, we think that the choice of 10 Ry cutoff is a reasonable trade-off between performance and computational effort.
[113] DLTS rates contain information on the energy difference between the fully relaxed charge states. It is the latter energy difference that corresponds to the levels calculated and discussed in this work. Measurements that use optical detection (recharging of the defect by optical means) access the charge transition energy without the contribution from the lattice relaxation in response to recharging. See also the remark in Ref. [43].
[114] J. Coutinho (private communication).
[115] S. B. Zhang, S.-H. Wei, and A. Zunger, Phys. Rev. Lett. **84**, 1232 (2000).
[116] S. Lany and A. Zunger, Phys. Rev. Lett. **98**, 045501 (2007).
[117] P. Rinke, A. Janotti, M. Scheffler, and C. G. Van de Walle, Phys. Rev. Lett. **102**, 026402 (2009).
[118] L. Heidin, Phys. Rev. **139**, A796 (1965).
[119] M. S. Hybertsen and S. G. Louie, Phys. Rev. Lett **55**, 1418 (1985).
[120] X. Qian, P. Umari, and N. Marzari, Phys. Rev. B **84**, 075103 (2011).
[121] J. Muscat, A. Wander, and N. M. Harrison, Chem. Phys. Lett. **342**, 397 (2001)
[122] B. G. Janesko, T. M. Henderson, and G. E. Scuseria, Phys. Chem. Chem. Phys. **11**, 443 (2009).
[123] A. Alkauskas, P. Broqvist, and A. Pasquarello, Phys. Rev. Lett. **101**, 046405 (2008).
[124] A. Alkauskas, P. Broqvist, and A. Pasquarello, Phys. Status Solidi B **248**, 775 (2011).
[125] S. Mukhopadhyay and N. M. Harrison, J. Magn. Magn. Mater. **324**, 3748 (2012).
[126] A. M. Rappe, K. M. Rabe, E. Kaxiras, and J. D. Joannopoulos, Phys. Rev. B **41**, 1227 (1990); **44**, 13175(E) (1991).
[127] G. Schwarz, Ph.D. thesis, Technical University of Berlin, 2002.
[128] J. Coutinho, R. Jones, V. J. B. Torres, M. Barroso, S. Öberg, and P. R. Briddon, J. Phys. Condens. Matter **17**, L521 (2005).